Towards the Interpretability Logic of all Reasonable Arithmetical Theories

Joost J. Joosten

12 december 1998

Abstract

The subject of this paper is the interpretability logic of all reasonable arithmetical theories, which we baptize *GIL*. This logic has long been conjectured by Albert Visser to be ILW^*. No modal completeness result was known for ILW^*. In order to provide a completeness proof of ILW^*, a good understanding of the sublogic ILM_0 seems indispensable. But there was no modal completeness result for ILM_0 either. In this paper a method for constructing models is developed. By this model-construction method we obtain a new proof of the modal completeness and decidability of *ILM*. Moreover do we obtain the modal completeness of ILM_0. Furthermore, a new principle, P_0, is introduced and studied. This principle is seen to be arithmetically valid and is completely independent with regard to the other principles studied here. The new logic ILP_0 turns out to be modally incomplete. The conjecture of ILW^* being the interpretability logic of all reasonable arithmetical theories is thus falsified.

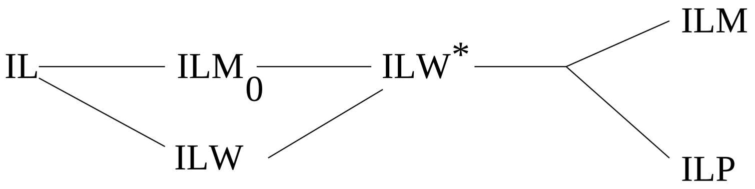
1 Introduction

A miracle happens. This is the openings phrase of an article by Albert Visser [Vis97] and hereby of this paper as well. The miracle mentioned emerges in metamathematical considerations of interpretations. Interpretations are interesting mathematical objects on their own. They frequently turn up in various areas. As a classical example one can mention the use of interpretations in establishing the undecidability of certain theories as exposed in [TMR53] Tarski, Mostowski, Robinson. The notion is frequently encountered in relative consistency results as well. Interpretations are also a very natural device when comparing different structures. Rather then paying importance to the names of the individual objects one would merely like to be able to talk about alikeness and imbeddability or better: interpretability. It turns out that the notion of interpretability can be formalized within the language of sufficiently strong mathematical theories. The methods used here are the same as when it came to formalizing the provability predicate. Actually, one makes explicit use of the formalized provability predicate to formalize interpretability and indeed both topics are closely related. Proceeding along like this yields a logic of interpretability with a Kripke-like semantics. As complex as is the notion of interpretability, as simple a method can be distilled representing this originally very complex notion. Solovay's method establishing arithmetical completeness of provability logic can be generalized to obtain arithmetically complete systems of interpretability logic and behold: decidability! Yes miracles do exist.

The logic of interpretability can thus be viewed as an extension to the logic of provability. The logic of provability includes important results concerning formal provability and provides a very sophisticated way of representing important results of, for example, Gödel, [Göd92], and Löb, [Löb55]. The subject of study in provability logic as with interpretability logic is formal theories. Provability is quite a stable notion though and does not distinguish between even very different theories. From quite weak theories onwards, all the provability logics come out the same. It is known that all theories in which $I\Delta_0 + \mathrm{EXP}$ is interpretable and which are Σ_1-sound under the translation, have the same provability logic: Löb's logic. See for example [Vis97], or [Vis84]. One could think of three approaches if one wants to obtain different logics, for fine tuning so to say.

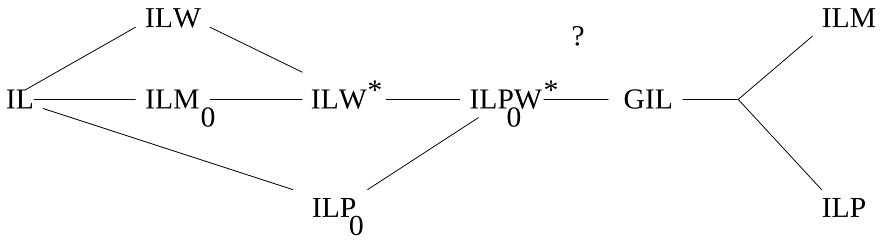
First one could try to descend to theories weaker than $I\Delta_0 + \mathrm{EXP}$, where Löbs logic is still arithmetically valid. This becomes an extremely difficult venture and despite intensive investigations very little is known on this subject. See e.g. [BV93]. Another direction would be to alter the bare logic. One could switch to, for example, intuitionistic logic in its relation to Heyting arithmetic obtaining a variety of new valid principles. Significant progress is made, [Iem98], [Vis94], but for example it is still unknown whether the provability logic of Heyting Arithmetic is axiomatizable at all. A third possibility would be to enrich the modal language. Where the logic of provability employs only one modal operator $\Box$, for formal provability, the language of interpretability logic includes also a binary modality $\rhd$ for interpretability. Although $\Box$ can be defined in terms of $\rhd$, one prefers to use both modalities. The whole provability logic now becomes a sublogic of the logic of interpretability. Moreover, distinctions between different theories are reflected by having different corresponding interpretability logics. So, interpretability is not that stable a notion as provability is. Furthermore, all sorts of arithmetical results can now be expressed, as we will see later, very elegantly like for example the model existence lemma, $\Diamond A \rhd A$, or a more intricate formalization of the second incompleteness theorem of Gödel, $\Diamond A \to \neg(A \rhd \Diamond A)$.

For two main interpretability logics, arithmetical completeness results are known. On the one hand that is the interpretability logic of *PA*, *ILM*, and on the other hand the interpretability logic of any (sufficiently strong) finitely axiomatizable theory, *ILP*. As these two logics are different, it seems very natural to ask for the core logic, that is the interpretability logic of all reasonable arithmetical theories. "Reasonable" in this context is not such a tight notion and can be read here as "containing $I\Delta_0 + \mathrm{SUPEXP}$". Looking for this notion is not just a matter of a simple intersection of all logics; Stronger logics can prove more but also have more expressive power. The primary aim of this paper is to contribute to the quest for the interpretability logic of all reasonable arithmetical theories. In due time many seemingly relevant interpretability logics have seen the light and many of them faded away already. The situation at the start of this research is sketched below.



IL, *ILP* and *ILM* are known to be decidable and modally complete. *ILM* and *ILP* are also arithmetically complete with regard to some theories. See e.g. [JdJ98]. In an article soon to appear, [dJV], *ILW* is shown to be decidable and modally complete. In [Vis90] *ILW* was put forward as the interpretability logic of all reasonable arithmetical theories. After the discovery that M_0 is a valid principle in [Vis91], the conjecture was updated to *ILW**($= ILWM_0$).

By the end of this paper we obtain a new picture which is given below. We develop a strategy for obtaining modal completeness without necessarily also obtaining decidability. This method is applied to *ILM* to get both completeness and decidability. The same method shows to be fruitful when it is applied to ILM_0; we obtain the modal completeness of ILM_0. ILW^* is shown to be inadequate as a candidate for *GIL*, the logic of all reasonable arithmetical theories. This is done by introducing a new arithmetically valid principle P_0, which is not derivable over ILW^*.



2 The landscape: Interpretability

2.1 The formalization of interpretability

The subject of our study as mentioned before is formal theories. The main tool used will be the notion of interpretability. There are many notions of interpretability, so we have to choose one for the default notion. When we talk in this paper about interpretability we refer to the notion of relative interpretability used in [TMR53]. A relative interpretation comprises a translation t and a relativizing formula $\sigma(x)$. t is a translation from the language of T' into the language of T. To every predicate constant P in the language of T', the translation assigns a formula P^t in the language of T. The relativizing formula σ is in the language of T as well. The translation is defined such that:

- $t(x = y)$ is $x = y$,
- for any other atom one has $t(P(\vec{x}))$ to be $P^t(\vec{x})$,
- t commutes with the Boolean connectives,
- $t(\forall x\alpha)$ is $\forall x(\sigma(x) \to t(\alpha))$ and consequently $t(\exists x\alpha)$ will be $\exists x(\sigma(x) \wedge t(\alpha))$.

So it will be clear what it means to say that a theory S interprets a theory T, namely S proves every translated theorem of T for some translation. This will be abbreviated by stating $S \rhd T$. Now if a theory T is strong enough one can do sufficient coding and hence talk within the theory T itself about interpretability. So within a theory one can formalize statements like $\alpha \rhd \beta$, stating "$T + \alpha$ interprets $T + \beta$". The intuitive reading should be something like "for some translation, $T + \alpha$ proves the translation of every theorem of $T + \beta$". And actually like this it can be formalized. For an extensive treatment one can see for example [JdJ98], [Vis97] or [Ber90]. If we are not too concerned about correct notation, we can think of formalized interpretability as the following Σ^0_3-sentence:

$$\alpha \rhd \beta \Leftrightarrow \exists J(\Box_T(\alpha \to \beta^J) \wedge \forall y(Ax_T(y) \to \Box_T(\alpha \to y^J))). \qquad (+)$$

In this sentence Ax_T is the formula expressing the fact: "y is the code of an axiom of T". The $\Box_T$ is a notion expressing provability in T. We will always use an intensional translation of the notion of provability. That is, not just any notion which externally happens to be the same as provability, but a notion in which the meaning of provability is really coded. See for example [Boo93]. It is only important that the provability notion will generate the Hilbert-Bernays conditions (also called Löb conditions) and consequently all of Löb's provability logic. There are provability predicates known for which the Hilbert-Bernays conditions are not derivable and for which the second incompleteness theorem of Gödel does not hold (see for example [Sha94] or [Vis89]), but we will not consider them. The J in (+) is the interpretation itself, comprising a translation from the one language into the other as well as a relativizing formula, defining the domain of the interpreted theory. If $T \rhd S$ is true, one has a uniform method of finding a model of S in any model of T. Thinking in terms of submodels can be very useful for acquainting oneself with some intuition on arithmetically valid principles. For example $S \rhd R \wedge R \rhd T \to S \rhd T$ can be thought of as a reflection of the following argument: "If in any model of S one can define a model of R, and if in any model of R one can define a model of T, one can consequently define in any model of S, a model of T."

2.2 Valid arithmetical principles

Some principles are easily seen to hold in a general arithmetical setting. From now on we will only study arithmetical theories that are reasonable. As mentioned before, reasonable can be read as "containing $I\Delta_0 + \mathrm{SUPEXP}$". We will treat some principles which hold in every reasonable arithmetical theory. Precisely these principles are later collected to be studied in a modal setting.

- We call $J1$ the principle of $\Box(\alpha \to \beta) \to \alpha \rhd \beta$. This reflects the fact that the identity function is a special case of an interpretation. If one takes J in (+) to be the identity, a tautology arises. (The relativizing formula can in this example just be taken $x = x$.)

- The principle $J2$ reads $\alpha \rhd \beta \wedge \beta \rhd \gamma \to \alpha \rhd \gamma$. We already encountered a plea for this principle when we viewed it as a statement of models. By more elementary means one can also see this principle to be true. The observation that the composition of two interpretations is again an interpretation, more or less explains the principle. Finally one should convince oneself that this reasoning can be done within the theory. (Here one has to use technical facts like $\vdash \Box(\beta \to \gamma^I) \to \Box(\beta^J \to \gamma^{J \circ I})$.)

- The principle $J3$ is $\alpha \rhd \gamma \wedge \beta \rhd \gamma \to \alpha \vee \beta \rhd \gamma$. Reading this as a statement of models legitimizes it immediately. If in any model of A, and in any model of B, one can define a submodel for C, well then in any model of $A \vee B$ one can define a submodel for C. For if $A \vee B$ holds, either one of them holds, but in both cases a model of C can be defined. In terms of interpretations $J3$ reflects that one can choose which interpretation to use, depending on α or $\neg\alpha$ to hold.

- Another principle, $J4$: $\alpha \rhd \beta \to (\Diamond\alpha \to \Diamond\beta)$, reflects that interpretations yield relative consistency results. One can also see $J4$ to be a direct consequence of $J1$ and $J2$. For this we write the consequent as $\Box\neg\beta \to \Box\neg\alpha$. Now suppose $\alpha \rhd \beta$ and $\Box\neg\beta$. $J1$ gives that $\Box\neg\beta = \Box(\beta \to \bot)$ implies $\beta \rhd \bot$. By transitivity , ($J2$), one gets $\alpha \rhd \bot$. But as the translation of $\bot$ under any interpretation is always $\bot$, we automatically get $\Box\neg\alpha$.

- The principle $J5$ reads $\Diamond\alpha \rhd \alpha$. It can be seen as an arithmetized version of the completeness theorem. The consistency of a statement implies the existence of a model on which this statement holds. As the theories we consider are strong enough to encode the whole Henkin construction, one can prove within that very theory $T + Con_T(\alpha) \rhd T + \alpha$. This coding can be done in a really weak theory, $T_0 \subseteq T$, so actually something stronger can be proven as well: $T_0 + Con_T(\alpha) \rhd T + \alpha$. This strengthening of $J5$ is essentially used in establishing the arithmetical validity of the new principle P_0 in paragraph 6.5.

All these principles are collected together in a modal logic. This modal logic is properly defined in chapter 3. It will be referred to as IL.

2.3 The other principles

If the theory under consideration is PA with full induction, more principles hold. We consider Montagna's principle $M: \ \alpha \rhd \beta \to \alpha \wedge \sigma \rhd \beta \wedge \sigma$ for every $\sigma \in \Sigma_1^0$. This can be seen to hold for PA using the following fact. Over PA one knows that $\alpha \rhd \beta$ iff every PA-model of α has an end extension which is a PA-model of β. So if the Σ_1^0-sentence σ holds in a PA-model M of α, there must be a witness in this model for σ. This witness serves as a witness as well in any end extension of M. We say that σ is being preserved under taking end extensions. So there exists an end extension for $\beta \wedge \sigma$, hence $\alpha \wedge \sigma \rhd \beta \wedge \sigma$.

If T, the theory under consideration, is finitely axiomatizable, the situation gets a lot easier. The sentence $\forall y(Ax_T(y) \to \Box(\alpha \to y^J))$ can just be replaced by $\Box(\alpha \to \tau^J)$, where τ is the conjunction of all the axioms of T. Consequently the sentence (+), expressing formal interpretability, becomes a Σ_1^0-sentence and hence can be "boxed up", using so-called provable Σ_1^0-completeness already used by Gödel. By doing so one obtains the so-called persistence principle $P: \ \alpha \rhd \beta \to \Box(\alpha \rhd \beta)$. As $I\Sigma_n$ is finitely axiomatizable for all $n \in \omega$, we have that P is an interpretability principle for $I\Sigma_n$ for all $n \in \omega$.

3 The modal logic of interpretability

3.1 The basic interpretability logic

The language of the modal interpretability logic is the language of provability logic extended with a binary modality $\rhd$. So we have at our disposal an enumerable supply of propositional variables, a unary modality $\Box$ and a binary modality $\rhd$. We can either use all the Boolean connectives or just constrain ourselves to $\wedge$ and $\neg$. We adhere to some reading conventions as to omit parentheses without introducing ambiguities. The negation, the diamond and the box bind stronger than $\wedge$ and $\vee$, which in turn bind stronger then $\rhd$ and $\to$. Finally we say that $\rhd$ binds stronger than $\to$. So $A \rhd B \to A \wedge \Box C \rhd B \wedge \Box C$, for example, is short for $(A \rhd B) \to ((A \wedge (\Box C) \rhd (B \wedge (\Box C))))$.

Definition 3.1 *The basic interpretability logic IL has the following axiom* *schemes:*

$$\begin{array}{ll}
L_1: & \Box(A \to B) \to (\Box A \to \Box B) \\
L_2: & \Box A \to \Box\Box A \\
L_3: & \Box(\Box A \to A) \to \Box A \\
J1: & \Box(A \to B) \to A \rhd B \\
J2: & A \rhd B \wedge B \rhd C \to A \rhd C \\
J3: & A \rhd C \wedge B \rhd C \to A \vee B \rhd C \\
J4: & A \rhd B \to (\Diamond A \to \Diamond B) \\
J5: & \Diamond A \rhd A
\end{array}$$

The rules of inference are modus ponens and necessitation, that is, if one has $\vdash A \to B$ and $\vdash A$, then also $\vdash B$, resp. $\vdash A \Rightarrow \vdash \Box A$.

Just as with provability logic, a Kripke-like semantics is given for *IL*.

Definition 3.2 *An IL-frame (also Veltman-frame) is a triple $(W, R, \{S_w \mid w \in W\})$ such that:*

1. *(W, R) is an L-frame, that is, W is a non empty set and R is a transitive conversely well-founded relation on W^2.*
2. *$S_w \subset w\uparrow \times w\uparrow$ $(w\uparrow := \{x \in W \mid wRx\})$.*
3. *$(R \cap w\uparrow) \subset S_w$.*
4. *S_w is reflexive.*
5. *S_w is transitive.*

By S we mean $\cup\{S_w \mid w \in W\}$. The elements of W will also be called *worlds* or *nodes.*

Definition 3.3 *An IL-model is an IL-frame together with a forcing relation $\Vdash$ between worlds and propositional letters. $\Vdash$ is extended to formulas by defining $\Vdash$ to commute with the Boolean connectives and defining*

- $w \Vdash \Box A \Leftrightarrow \forall w'(wRw' \to w' \Vdash A)$,
- $w \Vdash A \rhd B \Leftrightarrow \forall w'(wRw' \wedge w' \Vdash A \to \exists w''(w' S_w w'' \wedge w'' \Vdash B))$.

One could also first define $\Vdash$ and later determine the conditions on the binary relations S_w. In this manner one notices some frame correspondences. We see that $J4$ imposes 2, $J1$ corresponds with 4, $J2$ corresponds with 5 and $J5$ corresponds with 3. The axiom $J3$ has a special status as it does not impose anything on the frames.

Definition 3.4

- *$M, w \Vdash A$ and $w \Vdash A$ are equivalent expressions.*
- *$M \models A$ means that for all w in M, $M, w \Vdash A$.*
- *If F is a Veltman-frame, $F \models A$ means that for any $\Vdash$, $(F, \Vdash) \models A$.*
- *We say that A holds as a scheme on a frame F if all instantiations of A hold at F.*
- *Let $\mathcal{K}$ be a class of Veltman-frames. We define $\mathcal{K} \models A$ iff $\forall F \in \mathcal{K}\ F \models A$.*

As every axiom of *IL* is valid on every Veltman-frame, and this validity is preserved under the rules of inference, we see that *IL* is sound w.r.t. Veltman-frames. That is, all *IL*-derivable formulas hold on all Veltman-frames. The logic is modally complete as well. See [JV91]. So if a formula holds on all *IL*-frames, it must be derivable in *IL*. This completeness result is also presented for example in [JdJ98].

Throughout this paper we will use some very basic results of *IL*. Most of them are very easy to verify. In the next lemma we expose some facts of *IL*.

Lemma 3.5

- $IL \vdash A \rhd \bot \leftrightarrow \Box\neg A$
- $IL \vdash \Box\neg A \rightarrow A \rhd B$
- $IL \vdash A \vee \Diamond A \rhd A$
- $IL \vdash A \rhd A \wedge \Box\neg A$

These facts are indeed easy to verify as is partly done in [JdJ98]. Formal proofs in *IL* are quite laborious to write down. A Gentzen proof system is most likely hard to find.

3.2 More principles and frame correspondences

Other axioms we will consider are:

$$\begin{array}{lcl}
P & : & A \rhd B \to \Box(A \rhd B) \\
M & : & A \rhd B \to A \wedge \Box C \rhd B \wedge \Box C \\
W & : & A \rhd B \to A \rhd B \wedge \Box\neg A \\
M_0 & : & A \rhd B \to \Diamond A \wedge \Box C \rhd B \wedge \Box C \\
W^* & : & A \rhd B \to B \wedge \Box C \rhd B \wedge \Box C \wedge \Box\neg A \\
P_0 & : & A \rhd \Diamond B \to \Box(A \rhd B)
\end{array}$$

Definition 3.6 *The logic ILX is a modal logic in the language $\Box$, $\rhd$. All tautologies in this language are theorems of ILX. Further are all the axiom schemes of IL plus X itself, axiom schemes of ILX. The rules of inference are modus ponens and the rule of necessitation. X can be taken to be one of the above axiom schemes. The logic $ILXY$ is the logic where we add both X and Y as axiom schemes.*

The logic ILW^* is known to be precisely ILM_0W. See for example [Vis97]. All the logics we will consider over *IL*, will have the deduction theorem. The principle P_0 is introduced here for the first time. The principles P, M, etc., do not hold on every *IL*-frame. One can examine which condition the frame should satisfy in order to have the principle to hold. We obtain the following list of frame correspondences:

$$\begin{array}{lcl}
P & : & xRx'Ry \wedge yS_xy' \to yS_{x'}y'. \\
M & : & yS_xy' \wedge y'Rz \to yRz. \\
W & : & R \circ S \text{ is conversely well-founded.} \\
M_0 & : & xRx'Rx''S_xyRy' \to x'Ry'. \\
W^* & : & \text{Both the conditions of } W \text{ and of } M_0. \\
P_0 & : & xRx'Rx''S_xyRy' \to x''S_{x'}y'.
\end{array}$$

Instead of the formula one could read its universal closure. The frame condition of W is not first-order expressible though. The general problem of first order definability of a frame condition is known to be undecidable even in pure modal logic, and presumably even for extensions of Löbs logic, see Chagrova [Cha91], and is conjectured by van Benthem to be much worse. In [Ben84] he shows that the first-order definability for monadic second-order Π^1_1 sentences of a restricted form is non-arithmetically definable. The situation with interpretability logic is likely to be at least as complex. In a joint paper, Carlos Areces, Dick de Jongh and Eva Hoogland establish the interpolation property for *IL*, see [AdJH98]. As *IL* is a “nice” logic and as it has interpolation, they show, it must have the Beth property and hence fixed points; so any logic ILX has fixed points and thus, generalizing a result of Maksimova, also the Beth property.

3.3 Modal and arithmetical completeness results

Definition 3.7 *By the class of ILX-frames we mean the class of IL-frames where X hold as a scheme. This class is often also referred to as the class of characteristical ILX-frames.*

The Veltman semantics yields a uniform soundness theorem, that is, all *ILX* are sound w.r.t. their corresponding class of characteristical frames. So every derivable formula holds on every characteristical frame of that logic. One can ask if the reverse also holds. So if a formula holds on all *ILX*-frames, is it then automatically provable in *ILX*? No uniform method is known to settle this modal completeness problem and with every new logic a new proof has to be found. The logics ILM, ILP and ILW are known to be modally complete. The completeness proof of ILW of Dick de Jongh and Frank Veltman is to be published soon ([dJV]). For a long time the modal completeness of ILM_0 has been an open problem. In this thesis the question is answered positively. All modal completeness results in provability logic look alike a little. One works with maximal consistent subsets of a certain adequate set. With these maximal consistent sets as construction material, one makes up a countermodel for some statement $\mathcal{A}$ which is not provable in the logic. With any new logic, one has to re-consider the notion of adequateness. Adequate means large enough to do the required mathematics and small enough to have the countermodel finite. So in one stroke one proves modal completeness and decidability. (That is, if the logic is r.e. axiomatizable.) Miraculously this method work for most logics in this area. In this paper a method is developed by which only completeness is established without necessarily decidability.

Definition 3.8 *Given ILX. An* arithmetical translation *is a function $*$, assigning arithmetical formulas to formulas in the language $\Box$, $\rhd$ in such a way that $*$ commutes with all the connectives and moreover:*

- $(\bot)^* = \bot$,

- $(\Box A)^* =$ *an intensional formalization of "A^* is provable in ILX", for example the standard $\exists x Proof(x, \ulcorner A^* \urcorner)$,*

- $(A \rhd B)^* =$ *an intensional formalization of "ILX*$+A^*$ *interprets ILX*$+$ B^**". One can use* $(+)$ *for this purpose.*

Once this translation is introduced, one can ask for the logic of all principles proved by PA under this translation. In 2.2 and 2.3 we have already seen all the principles of *ILM* to be provable under the translation $*$. But something stronger holds. Solovay's first completeness theorem can be generalized to interpretability logic to obtain the arithmetical completeness of *ILM* w.r.t. PA.

Theorem 3.9 *(Berarducci [Ber90], Shavrukov [Sha88]) ILM* $\vdash A \Leftrightarrow \forall^* \; PA \vdash A^*$.

Actually *ILM* is the interpretability logic of any essentially reflexive theory. It turns out that also in this situation adding the reflexion principle is sufficient to obtain a generalization of Solovay's second completeness result. See [Ber90]. All principles of ILP have been seen to be arithmetically valid in every reasonable finitely axiomatizable arithmetical theory. We also have ILP arithmetically complete w.r.t. any such theory. So ILP is the interpretability logic of $I\Sigma_n$ for all $n \in \omega$ as every $I\Sigma_n$ is finitely axiomatizable.

We have a notion of reasonable arithmetical theory. For the time being this notion can be thought of as $I\Delta_0 +$ SUPEXP. Up until today it is unknown what the interpretability logic of all reasonable arithmetical theories is. For a long time it was conjectured to be *ILW*. See [Vis91]. Until M_0 was seen to be arithmetically valid. Also in [Vis91]. Since then *ILW*$^*(= ILWM_0)$ has been conjectured to be the interpretability logic of all reasonable arithmetical theories. From now on we will abbreviate this target logic by *GIL*, for general interpretability logic. We will not write ILG as we do not want to insinuate that we would know the relevant principle. A priori it is not even known if *GIL* is axiomatizable at all! In this paper the conjecture that *ILW** is *GIL* is falsified.

Definition 3.10 T' *is* Π^0_1*-conservative over* T *means that* $T' \vdash \pi$ *implies* $T \vdash \pi$ *for every* $\pi \in \Pi^0_1$.

The definition could be generalized by fixing different sets Γ of sentences instead of just Π^0_1. One can also translate $A \rhd B$ to an intensional formalization of "$T + B$ is Π^0_1-conservative over $T + A$". The logic corresponding to this translation is the logic of Π^0_1-conservativity. It is known that *ILM* is the logic of Π^0_1-conservativity for any arithmetical theory containing $I\Sigma_1$. If an arithmetical theory is essentially reflexive the notion of Π_1^0-conservativity coincides with the notion of interpretability.

3.4 The aim of this paper

We will present a method for establishing modal modal completeness results which uncouples the completeness result from the finite model property. This method is first applied to *ILM*. So a new proof of the modal completeness of *ILM* is provided. There is an option built in in this construction method to also obtain the finite model property and hence the decidability of a certain logic. By this option we obtain the decidability of *ILM* in chapter 4. This construction method is then applied in chapter 5 to obtain the completeness of ILM_0. In this case however, we did not succeed in proving decidability. In chapter 6 we present a new modal principle and use generalized Veltman semantics to subject it to a modal study.

4 The modal completeness and decidability of *ILM* via the construction method

4.1 The general construction of the proof

The general philosophy to a modal completeness result is always the same. In order to show the completeness of some logic L w.r.t. its class of characteristic frames $\mathcal{K}$ one has to prove that

$$\forall \mathrm{A}\ [\ \forall \mathcal{M} \in \mathcal{K}\ \ \mathcal{M} \models \mathrm{A} \longrightarrow L \vdash \mathrm{A}\ ]$$

or equivalently

$$\forall \mathrm{A}\ [\ L \nvdash \mathrm{A} \longrightarrow \exists \mathcal{M} \in \mathcal{K}\ \ \mathcal{M} \not\models \mathrm{A}\ ].$$

Thus the main approach is clear: given a modal sentence $\mathcal{A}$ that is not derivable in L, one has to provide a model $\mathcal{M}$ and a node m that forces $\neg\mathcal{A}$, that is, $\mathcal{M}, m \Vdash \neg\mathcal{A}$. Moreover one wants the frame of this model to be in the class of characteristic frames. We reserve the symbol $\mathcal{A}$ to designate a formula that is not derivable over *ILM*. (The font of the symbol $\mathcal{A}$ is a little unusual in the modal setting, but we want to have the A free to use in the course of this chapter.) In all completeness proofs the basic material for the construction of a countermodel consists of maximal consistent sets of sentences. So a world of such a countermodel will comprise a copy of a maximal consistent set of sentences. One combines these sets in such a way that eventually one obtains a so-called truth lemma, correlating membership of a sentence to a set, to the forcing of that sentence in that particular world/set. In our approach presented below we will do exactly this. Distinctions though can be found in the way the countermodel is provided and the material that is used.

In modal completeness results one often wants, in addition, to prove the decidability of that logic. If an r.e. axiomatizable logic has the finite model property it automatically is decidable. For this reason one always looks for a finite countermodel in the completeness proof. One essential ingredient for obtaining finiteness in provability logic is to not work with full maximal consistent sets of modal sentences, but with sufficiently large truncated parts of them instead. These truncated parts are maximal consistent subsets of a so-called adequate set of sentences. In choosing the adequate set one is driven by two opposite motives. One does not want the adequate set too large, because of the finiteness and hence the decidability. It is generally speaking also more difficult to match the consistent sets in a proper way if they contain more sentences. Nor does one wants the adequate set too small because one wants to have the truth lemma for sufficiently many sentences, which therefore must be in the set. Once the adequate set is chosen, the relations between the maximal consistent subsets of the adequate set are defined. By doing so, one obtains the so-called canonical model for which you prove the truth lemma.

In our approach we will not define in one blow the canonical model but instead we will inductively build up a model. Further we will adhere to the method of "adequate set abuse". This means that your adequate set is very large but we only want the truth lemma for a finite set of so-called *relevant sentences*. We also use a different finite set for ensuring that our R-relation is conversely well-founded. In the old method one single set had to take care of all the jobs that are here done by three different sets. Our model $\langle M, R, S, \Vdash\rangle$ will be built up in stages out of copies of maximal *ILM*-consistent sets of modal sentences in such a way that eventually the following truth lemma holds:

$$\forall \mathrm{A} \in R \ \ \forall m \in M \ \ [m \Vdash \mathrm{A} \Leftrightarrow \mathrm{A} \in m].$$

R is the set of *relevant sentences.* It is the minimal set containing $\neg\mathcal{A}$ that is closed under taking subformulas and single negation. Since $\mathit{ILM} \nvdash \mathcal{A}$ there is a maximal *ILM*-consistent set m_0 with $\neg\mathcal{A} \in m_0$. By building an *ILM*-model "containing" m_0 we obtain a countermodel to $\mathcal{A}$.
The proof of the truth lemma will run as follows:

- $\mathcal{M}, m \Vdash P \Leftrightarrow P \in m$ by definition of $\Vdash$ for propositional variables P.
- $\mathcal{M}, m \Vdash \neg \mathrm{A} \Leftrightarrow \mathcal{M}, m \nVdash \mathrm{A} \Leftrightarrow \mathrm{A} \notin m \Leftrightarrow \neg \mathrm{A} \in m$.
- $\mathcal{M}, m \Vdash \mathrm{A} \wedge \mathrm{B} \Leftrightarrow \mathcal{M}, m \Vdash \mathrm{A}$ and $\mathcal{M}, m \Vdash \mathrm{B} \Leftrightarrow \mathrm{A},\mathrm{B} \in m \Leftrightarrow \mathrm{A} \wedge \mathrm{B} \in m$.
- $\mathcal{M}, m \Vdash \mathrm{A} \rhd \mathrm{B} \Leftrightarrow \mathrm{A} \rhd \mathrm{B} \in m$ by the construction of $\mathcal{M}$. Actually the two directions are separately taken care of in the construction.
  "$\Leftarrow$" If $\mathrm{A} \rhd \mathrm{B} \in m$ then we will have constructed the model so that indeed $\mathcal{M}, m \Vdash \mathrm{A} \rhd \mathrm{B}$.
  "$\Rightarrow$" If $A \rhd B \notin m$ then by the maximality of m we have that $\neg(A \rhd B) \in m$. This is also a relevant formula. Therefore the construction is such that one has $\mathcal{M}, m \Vdash \neg(A \rhd B)$.

The countermodel must be an *ILM*-model. The M axiom states $A \rhd B \rightarrow A \wedge \Box C \rhd B \wedge \Box C$ and it demands the characteristic *ILM*-frames to satisfy the following condition: $xSyRz \rightarrow xRz$. Thus in the construction of our model we want that every possible R-successor of such a world y, can consistently be taken to be an R-successor of x. This can be done if we demand $x \subset_{\Box} y$. (This means $\Box A \in x \rightarrow \Box A \in y$.) We will prove a lemma that guarantees one can always incorporate this condition. It is now time to prepare for the construction and develop some tools.

4.2 Tools

Definition 4.1 *With a fixed sentence A we associate a corresponding set of* relevant sentences $R(A)$*, or sometimes written just as R. This is the smallest set of sentences containing $\neg A$ which is closed under taking single negation and subformulas.*

Definition 4.2 *We say $Prop(B)$ is the set of propositional variables occurring in B and define $\widetilde{M} := \{\Gamma \mid \Gamma$ maximal ILM-consistent set of modal sentences s.t. $\forall \mathrm{A} \in \Gamma, \quad Prop(\mathrm{A}) \subset Prop(\mathcal{A})\}$.*

Definition 4.3 *$BR(\mathcal{A})$ is defined to be the smallest set including the relevant sentences $R(\mathcal{A})$, such that if $A \rhd B \in R(\mathcal{A})$ then both $\Box\neg A$ and $\Box\neg B$ are in $BR(\mathcal{A})$.*

Definition 4.4 *We define for $\Gamma, \Delta \in \widetilde{M}$*
$\Gamma \prec \Delta \Leftrightarrow \forall \mathrm{A}\ (\ \Box \mathrm{A} \in \Gamma \Rightarrow \mathrm{A}, \Box \mathrm{A} \in \Delta)$ *and* $\exists \Box \mathrm{A} \in (\Delta \setminus \Gamma) \cap BR(\mathcal{A})$

Definition 4.5 *We define for* $\Gamma, \Delta \in \widetilde{M}$
$\Gamma \prec_B \Delta \Leftrightarrow \Gamma \prec \Delta$ *and for all* $\mathrm{A} \rhd \mathrm{B} \in \Gamma$ *one has* $\neg\mathrm{A}, \Box\neg\mathrm{A} \in \Delta$. *We say that* Δ *is a* B-critical *successor of* Γ.

The following three lemmas form the mathematical fundaments of the modal completeness proof of ILM.

Lemma 4.6 *Let* $\Gamma \in \widetilde{M}$ *and* $\neg(A \rhd B) \in \Gamma \cap R$. *There exists* Δ *such that* $\Gamma \prec_B \Delta$ *and* $A \in \Delta$. *Moreover,* Δ *can be chosen to be maximal w.r.t. the number of* $\Box$*-formula‘s.*

Proof of Lemma 4.6. It is good to note the clause concerning the maximality of Δ. This is a new subtlety and has been added to enable us to keep the model finite. It will prevent us from endlessly having to repeat actions because at some stage we can be sure that a previously obtained maximal consistent set will do the job we want it to do. The same holds for lemma 4.8. So, let Γ satisfy the conditions of the lemma. The following set

$$\{C, \Box C \mid \Box C \in \Gamma\} \cup \{\neg D, \Box\neg D \mid D \rhd B \in \Gamma\} \cup \{A, \Box\neg A\}$$

is ILM-consistent. For suppose it were not, then

$$\{C, \Box C \mid \Box C \in \Gamma\} \cup \{\neg D, \Box\neg D \mid D \rhd B \in \Gamma\} \cup \{A, \Box\neg A\} \vdash \bot$$

($\vdash$ means $\vdash_{ILM}$ in this setting.) Then, by compactness, for some finite selection of sentences one obtains:

$$\begin{array}{ll}
C_1, \ldots, C_n, \Box C_1, \ldots, \Box C_n, \neg D_1, \ldots, \neg D_m, \Box\neg D_1, \ldots, \Box\neg D_m, A, \Box\neg A \vdash \bot & \Leftrightarrow \\
C_1, \ldots, C_n, \Box C_1, \ldots, \Box C_n, A, \Box\neg A \vdash (\bigwedge_{i=1}^m \neg D_i \wedge \bigwedge_{i=1}^m \Box\neg D_i) \to \bot & \Leftrightarrow \\
C_1, \ldots, C_n, \Box C_1, \ldots, \Box C_n, A, \Box\neg A \vdash \bigvee_{i=1}^m D_i \vee \bigvee_{i=1}^m \Diamond D_i & \Leftrightarrow \\
C_1, \ldots, C_n, \Box C_1, \ldots, \Box C_n \vdash A \wedge \Box\neg A \to \bigvee_{i=1}^m D_i \vee \bigvee_{i=1}^m \Diamond D_i & \Leftrightarrow \\
\Box C_1, \ldots, \Box C_n \vdash \Box(A \wedge \Box\neg A \to \bigvee_{i=1}^m D_i \vee \bigvee_{i=1}^m \Diamond D_i) & \Leftrightarrow \\
\Box C_1, \ldots, \Box C_n \vdash A \wedge \Box\neg A \rhd \bigvee_{i=1}^m D_i \vee \bigvee_{i=1}^m \Diamond D_i & \Leftrightarrow \\
\Box C_1, \ldots, \Box C_n \vdash A \rhd \bigvee_{i=1}^m D_i \vee \bigvee_{i=1}^m \Diamond D_i &
\end{array}$$

But as for each $i : D_i \rhd B \in \Gamma$, one obtains

$$\Gamma \vdash \bigvee_{i=1}^{m} D_i \vee \bigvee_{i=1}^{m} \Diamond D_i \rhd B$$

hence $\Gamma \vdash A \rhd B$. But this would imply $A \rhd B \in \Gamma$, quod non.
It has thus been shown that the former set of sentences is ILM-consistent

and hence it is included in some element $\Delta' \in \widetilde{M}$. Note that $\Box\neg A \in (\Delta' \setminus \Gamma) \cap BR(\mathcal{A})$, so that really one has $\Gamma \prec \Delta'$. Now Δ' can be chosen to be maximal with respect to $\Box$-inclusion. For if one has a chain

$$\Delta' = \Delta_0 \subset_\Box \Delta_1 \subset_\Box \Delta_2 \subset_\Box \dots \ (\ \Delta_i \subset_\Box \Delta_j \text{ means } \Box A \in \Delta_i \Rightarrow \Box A \in \Delta_j),$$

with $\Gamma \prec_{\mathrm{B}} \Delta_i$, one can consider $\Delta_\infty := \bigcup_{i=1}^{\infty}\{\Box A \mid \Box A \in \Delta_i\}$.

Clearly $\{A\} \cup \Delta_\infty \cup \{\neg C, \Box\neg C \mid C \rhd B\} \cup \{D \mid \Box D \in \Gamma\}$ is now a consistent set. For suppose it were not, then

$$A, \Delta_\infty, \{\neg C, \Box\neg C \mid C \rhd B\}, \{D \mid \Box D \in \Gamma\} \vdash \bot$$

and by compactness you see that Δ_∞ can be replaced by some finite part. But this finite part of Δ_∞ is then contained in some Δ_{i_0}. However since $A \in \Delta_{i_0}$ and $\{\neg C, \Box\neg C \mid C \rhd B\} \cup \{D \mid \Box D \in \Gamma\} \subset \Delta_{i_0}$ we would have that $\Delta_{i_0} \vdash \bot$, which is not the case. So $\{A\} \cup \Delta_\infty \cup \{\neg C, \Box\neg C \mid C \rhd B\} \cup \{D \mid \Box D \in \Gamma\}$ is consistent and can thus be extended to a maximal consistent set which is an upper bound of the chain. By Zorn's Lemma we get a maximal element $\Delta\ {}_{B}{\succ}\ \Gamma$.

QED

Lemma 4.7 *Let $\Gamma \in \widetilde{M}$ with $A \rhd B \in \Gamma$ and let $\Delta \in \widetilde{M}$ be such that $\Gamma \prec_C \Delta$ and $A \in \Delta$. There exists $\Delta'\ {}_{C}{\succ}\ \Gamma$, with $B \in \Delta'$.*

PROOF OF LEMMA 4.7. The proof of this lemma is quite similar to that of lemma 4.6. Suppose A, B, C, Γ, and Δ satisfy the assumptions of the lemma, and suppose there were no such Δ'. It would then follow again by compactness that there exist $\Box D_1, \dots, \Box D_m \in \Gamma$ and $E_1 \rhd C, \dots, E_n \rhd C \in \Gamma$ such that

$$D_1, \dots, D_m, \Box D_1, \dots, \Box D_m, \neg E_1, \dots, \neg E_n, \Box\neg E_1, \dots, \Box\neg E_n, A, \Box\neg A \vdash \bot$$

By reasoning completely analogously as in the proof of lemma 4.6 one again obtains $\Gamma \vdash B \rhd C$ and therefore also $B \rhd C \in \Gamma$. By transitivity we have $A \rhd C \in \Gamma$ as well. In view of the fact that Δ is supposed to be a C-critical successor of Γ, it should hold that $\neg A$ as well as $\Box\neg A$ are in Δ. This is clearly contradictory to the fact that A is already in Δ. The existence of the required Δ' is thereby demonstrated.

QED

Lemma 4.8 *Let $\Gamma \in \widetilde{M}$ with $A \rhd B \in \Gamma$ and let $\Delta \in \widetilde{M}$ be such that $\Gamma \prec_C \Delta$ and $A \in \Delta$. There exists $\Delta'\ {}_{C}{\succ}\ \Gamma$, $B \in \Delta'$ and moreover*
$\Box E \in \Delta \Rightarrow \Box E \in \Delta'$. Again Δ' can be chosen to be maximal with respect to $\Box$-inclusion.

PROOF OF LEMMA 4.8. We again suppose the lemma to be false and note that therefore the set

$$\{D, \Box D \mid \Box D \in \Gamma\} \cup \{\Box E \mid \Box E \in \Delta\} \cup \{\neg F, \Box\neg F \mid F \rhd C \in \Gamma\} \cup \{A, \Box\neg A\}$$

must be inconsistent. And thus, by compactness, one obtains for some finite selection of sentences

$$\begin{array}{l} D_1, \ldots, D_l, \Box D_1, \ldots, \Box D_l, \Box E_1, \ldots, \Box E_m, \\ \neg F_1, \ldots, \neg F_n, \Box\neg F_1, \ldots, \Box\neg F_n, A, \Box\neg A \qquad \vdash \quad \bot \ . \end{array} \tag{1}$$

Γ is a maximal *ILM*-consistent set. So, as $A \rhd B \in \Gamma$, one also has $A \wedge \bigwedge_{i=1}^{m} \Box E_i \rhd B \wedge \bigwedge_{i=1}^{m} \Box E_i \in \Gamma$. Because A and E_i for $i = 1, \ldots, m$ are all in Δ, also $A \wedge \bigwedge_{i=1}^{m} \Box E_i$ is in Δ . Now it is possible to apply lemma 4.7 to $A \wedge \bigwedge_{i=1}^{m} \Box E_i \rhd B \wedge \bigwedge_{i=1}^{m} \Box E_i$. This yields a C-critical successor Δ' of Γ with $B \wedge \bigwedge_{i=1}^{m} \Box E_i \in \Delta'$. The consistency of Δ' conflicts with (1), because all the premises of the derived contradiction are in Δ'. The argument showing that the Δ' can be chosen to be maximal with respect to the $\Box$-inclusion is completely the same as in lemma 4.6.

QED

4.3 The construction

We have now provided ourselves with enough tools to start with the actual construction. Our construction material will consist of copies of elements of the earlier defined set $\widetilde{M}$. Step by step we will paste them together by means of defining the R and the S relation. So, again consider an $\mathcal{A}$ for which we have $ILM \nvdash \mathcal{A}$. Since $\mathcal{A}$ is not provable in *ILM*, there exists an m_0 in $\widetilde{M}$ containing $\neg\mathcal{A}$. This m_0 will be the starting point of the construction.

In order to be able to talk about the model under construction it is convenient to first introduce some ad hoc nomenclature.

Definition 4.9 *For any world x of the model we define the* order *of this world to be the number of boxed formulas in $x \cap BR(\mathcal{A})$ minus the number of boxed formulas in $m_0 \cap BR(\mathcal{A})$. A boxed formula is a formula of the form $\Box A$.*

Definition 4.10 *A* problem *in x is a relevant sentence $\neg(A \rhd B) \in x$ such that there is no y with $xR_e^B y$ and $A \in y$. The R_e^B relation is just a special case of the R relation and will be introduced and treated later on. When we talk of the order of some problem we mean the order of the world in which the problem "lives".*

Definition 4.11 *A* deficiency *in x w.r.t. y is a relevant sentence $C \rhd D \in x$ such that $C \in y$ and xRy but there is no y' with yS_xy' and $D \in y$.*

Definition 4.12 *For any world x in the model we define C^B_x, the B-*critical cone *of x, to be the smallest set of worlds such that $xR^B_ey \to y \in C^B_x$ and $y \in C^B_x \wedge yS_xy' \to y' \in C^B_x$.*

The model will be built up in stages. At every stage an *ILM*-model is made out of the *ILM*-model of the previous stage. Eventually a model is obtained in which the desired truth lemma will hold. For example the R relation is not defined all at once but will be expanded as the construction proceeds. The only entity that will be globally defined is the $\Vdash$ relation. It will be defined as in all modal completeness proofs: $x \Vdash P$ iff $P \in x$ for propositional variables P. (Of course, we somewhat sloppily write $A \in x$ if A belongs to $\tilde{m}$ of which x is a copy.) The construction method can schematically be represented by the following:

- The first approach to our final model will be to set $M = \{m_0\}$, $R = \varnothing$ and $S = \varnothing$. Then enter the following loop:

begin
As long as problems still exist in the model, execute the following two steps:

- Locate a problem of minimal order and eliminate it. This should be done in such a way that no new problems of the same or lower order will be created. Close off under the *ILM* frame conditions in a minimal way. (In order to have the *ILM* frame condition, $R \subset S$, transitivity of both R and S, etcetera.)

- By having eliminated one problem, probably loads of deficiencies have emerged as a side-effect. Eliminate all these deficiencies. Again the resolving of these deficiencies should not create new problems of the same or lower order.

end

If termination of this process can be established then it is clear that the truth lemma holds. For the only part of the truth lemma which involves $\rhd$ and hence the only part of the truth lemma that needs a proof could be formulated as "there are no problems nor deficiencies". After the initialization of the procedure, no deficiencies exist because in the model with just one node and no relations also $\Box\bot$ holds. At the end of each loop no deficiencies exist either. So if the process terminates there will be no more deficiencies. But if the process terminates this means that all problems have been eliminated. Now if the truth lemma holds, the observation that $\neg\mathcal{A} \in m_0$ concludes the completeness proof.

So it remains to show that both problems and deficiencies can be eliminated in such a way that termination is guaranteed.

4.3.1 Problems

Easiest to deal with are the problems. Say that $\neg(A \rhd B)$ is a problem for some world m. To eliminate this problem is to provide this m with an R-successor m' where A holds, in such a way that from m' it is impossible to reach some world where B holds via an S_m-transition. As B should not hold in m' nor at any possible R-successor of m', we should take m' to lie B critically above m. We also must take care that starting in m' with an S_m-transition, it will never be possible to reach a world where B holds. We should so to say "fence the S_m-scope of m' in". This fencing in is performed by the so-called *B-critical cone of m*. We write C_m^B. So the whole B-critical cone of m will lie B-critically above m, and the definition will insure that it will not be possible to leave the B-critical cone of m with an R or an S_m transition. By doing so, we are sure that this particular problem $\neg(A \rhd B) \in m$ will never re-emerge.

The existence of a B-critical successor of m containing A is guaranteed by lemma 4.6. In our construction we thus define this entity m' to be an R_e successor of m. We write R_e instead of R because we want to be able to distinguish *essential R relations* which are added to eliminate a problem, from *non-essential R relations* (we write R_n) which are added either to restore the *ILM* frame conditions or to eliminate deficiencies. By R we mean the transitive closure of $R_e \cup R_n$. Sometimes we write R_e^B to indicate that the added R_e transition is an intended B-critical one. It will turn out to be useful to have the maximality w.r.t. $\Box$-inclusion of m'. By this the finiteness of the construction can be guaranteed. All the R-successors of this m' will automatically be B-critical successors of m. It is therefore sufficient to ensure that by an S_m-transition it is impossible to leave the B-critical cone of m. This is incorporated in the construction and expressed by incariant 3 below.

4.3.2 Deficiencies

Deficiencies are harder to deal with and they demand an inductive treatment. Eliminating a problem was done by just defining some R successor. In this part we need to define the S transitions. They will be defined in such a way that one always has

$$yS_xz \to (\Box A \in y \to \Box A \in z) \quad (*)$$

By doing so, the special M frame condition can readily be incorporated. Actually $(*)$ should properly be verified. One way of doing so is by means of **invariants**. The construction can roughly be seen as an initialization followed by a loop. An invariant is a property which holds after the initialization and after every execution of the loop. One of the invariants has already been encountered: "there are no deficiencies". Other useful invariants are:

1. The model is a finite ILM-model.
2. $xRy \Rightarrow x \prec y$.
3. $y \in C_x^B \to x \prec_{\mathrm{B}} y$.
4. $B \neq B' \to C_x^B \cap C_x^{B'} = \varnothing$.
5. $ord(x) = ord(y) \wedge x \neq y \to C_x^B \cap C_y^{B'} = \varnothing$.

All these properties can easily be checked while going through the construction. It should be made clear that these invariants hold at the beginning and after every execution of the loop.

The problem $\neg(A \rhd B)$ in m was taken care of by defining a B-critical successor m' to be its R_e^B successor. The thus obtained extension is closed off under the frame conditions of ILM in a minimal way. By this we mean that the new S is taken to be the reflexive transitive closure of the previous one plus $\langle m, m' \rangle$, and the new R is taken to be the transitive closure of "new S" $\circ R$. This again yields an ILM-frame.

If there are deficiencies in m with respect to m' we will prove by induction on $ord(m)$ that all these deficiencies can be eliminated. More generally we prove by induction on $ord(x)$ that all deficiencies in x w.r.t. y can be eliminated without new deficiencies occurring.

- $ord(x) = 0$. This means $x = m_0$. It follows from the invariants that $y \in C^B_{m_0}$ for at most one B. Our task is to eliminate all deficiencies in m_0 w.r.t. y while preserving all invariants and in particular invariant 3. All deficiencies in x (in this case $x = m_0$) w.r.t. y are dealt with by the so-called process of making an S^y_x-block. It is described below how this proceeds precisely. The main idea of this process is to provide y with a whole net of S_x-successors insuring that no deficiencies exist in x w.r.t. y or w.r.t. any other element y' in the S^y_x-block. Formally speaking, the term S_x-successor is somewhat misleading, suggesting a successor relation. We will not worry too much about this formally wrong name. The whole block is created in such a way that it lies B-critically above x. By doing so, at least in this step, no old, already dealt with, problems will re-emerge. This is the general philosophy behind the process of making an S^y_x-block.

- $ord(x) = n + 1$. We could eliminate all the deficiencies in x w.r.t. y by again making an S^y_x-block. This however could create new deficiencies of order $n + 1$ in other worlds. For example if $x'Sx$. The typical M frame condition demands the whole S^y_x-block to be above x'. If also $ord(x') = n + 1$, we probably have loads of new deficiencies of order $n + 1$ by eliminating some in x. The induction hypothesis does not apply to them. For this reason not only the deficiencies in x are eliminated at this stage, but at the same time the deficiencies of all the worlds $x'Ry$ with $ord(x') = n + 1$. We call the set of all these worlds N^y_x. The procedure now consists of three parts.

  Because of invariant 5 one has that y can only be in some $C^B_{x'}$ for at most one x' in N^y_x. We call this world x_0, and start by making an $S^y_{x_0}$-block to resolve the deficiencies in x_0 w.r.t. y. By doing so we have eliminated all deficiencies in x_0 while conserving the B-criticality. So if $x_0 R^B_e y$, we make the whole $S^y_{x_0}$-block lie B-critically above x_0. If no such x_0 exists we just omit this part.

  For every $x' \in N^y_x$ we have to eliminate deficiencies if any, w.r.t. all elements of the $S^y_{x_0}$-block. The process of doing so is called the proces of making an S^y_N-block. This is discussed in more detail below. N is some set of worlds, wich in our case we take to be N^y_x. So for each y' in the $S^y_{x_0}$-block, an $S^{y'}_{N^y_{x_0}}$-block is made to eliminate all deficiencies in N^y_x w.r.t. that world y'. Note that $N^y_x = N^y_{x_0}$. If one leaves the

B-critical cone of x_0, this can only be done by an $S_{z \neq x_0}$-successor. So again the B-critical cones are respected. After having done all this there are no more deficiencies in N^y_x w.r.t. the just created blocks.

For $x' \neq x_0$ there might still be some deficiency w.r.t. y though. These are dealt with by making an $S^y_{N^y_{x'}}$-block. All deficiencies in N^y_x are thus eliminated. By doing so only new deficiencies of lower order may have arisen, but the induction hypothesis takes care of them.

4.3.3 The details of constructing the blocks

Making an S^y_x-block. "Making an S^y_x-block" is the process which is applied to eliminate all deficiencies in x w.r.t. y, when $y \in C^B_x$ for some B. Eliminating a deficiency in x w.r.t. y is nothing but providing an S_x-successor of y that suits the job. If we want to respect invariant 3 we must ensure that every S_x-successor of y is B-critical as well. It turns out to be possible to define a finite construction of S_x-successors of y such that no deficiency remains in x w.r.t. y and w.r.t. the whole S^y_x-block, and such that moreover this whole construction lies B-critically above x. This is our so-called process of "making an S^y_x-block", and it proceeds as follows.

- We first set the S^y_x-block to be empty.
- Now for every deficiency in x w.r.t. y we put a y' in the S^y_x-block which repairs the deficiency in a sufficient way. If $C \rhd D \in x$ is such a deficiency, with $C \in y$, lemma 4.8 guarantees the existence of a world y' such that $x \prec_{\mathrm{B}} y', D \in y', y \subset_\Box y'$, and y' being maximal in this $\subset_\Box$-ordering. Note that there is no ambiguity of the B since invariant 4 guarantees its uniqueness. If we define yS_xy', this y' has repaired the deficiency $C \rhd D \in x$. We have to add also xRy', etcetera. In other words we have to close under the *ILM* frame conditions in a minimal way. After all this no deficiencies in x exist any more w.r.t. y. There might be very well a whole lot of deficiencies in x w.r.t. the S^y_x-block constucted so far. They are taken care of in the following loop.

Repeat the following steps as long as deficiencies of x w.r.t. the S^y_x-block still do exist.

- Fix a y' in the S^y_x-block such that there is a deficiency of x w.r.t. y'. Say $C \rhd D$ is a deficiency of x w.r.t. y'. If there is a y'' in the S^y_x-block with $D \in y''$ and $\forall E(\Box E \in y' \to \Box E \in y'')$, then define $y' S_x y''$. If there is no such y'' then use lemma 4.8 to obtain a B-critical successor y'' of x with $D \in y''$ such that $\forall E(\Box E \in y' \to \Box E \in y'')$ and moreover such that y'' is maximal with respect to this $\Box$-inclusion. Again define $y' S_x y''$.

- Close off under the frame conditions of *ILM* in a minimal way.

It is evident that the process of making an S_x^y-block will terminate. There is only a finite number of possible deficiencies of x w.r.t. any specific y'. This implies that every world in the S_x^y-block needs only a finite number of S_x-successors. An arbitrary chain (without loops) of S_y-successors is limited in length due to the maximality of all worlds w.r.t. the $\Box$-inclusion. By the clause "If there is a y'' in the S_x^y-block with $\ldots$, we are forced to use the same y'' again if it is approppiate. These two ingredients ensure the finiteness of the S_x^y-block.

Making an S_N^y-block. The procedure of "making an S_N^y-block" is quite similar to that of "making an S_x^y-block". The main difference is that we are eliminating deficiencies here of a whole bunch of worlds at the same time. Another thing is that we are not too worried about criticalness. Of all the worlds in N, for at most one $x_0 \in N$, we have that $y \in C_{x_0}^B$ by invariant 5. The procedure of "making an S_N^y-block" is meant to deal with all deficiencies in $N \setminus \{x_0\}$ w.r.t. y in such a way that no new deficiencies in N w.r.t. the S_N^y-block occur. But where we do not take x_0 into account while eliminating the deficiencies in N w.r.t. y, we might need to take x_0 into account while eliminating the deficiencies in N w.r.t. the S_N^y-block. Invariant 3 is not violated though, because it is not possible to enter the S_N^y-block from y via an S_{x_0}-successor. Therefore if some worlds in S_N^y lie above x_0 (and this is possible if we eliminate a deficiency in x with $x_0 S x$), they will not be in the B-critical cone of x_0 and hence we will not need to worry about their B-criticality. Their successors however end up in the B-critical cone anyway via the *ILM*-condition ($S \circ R \subseteq R$). But this is alright, since, if $x' \prec_{\mathrm{B}} x \prec y'$ with $x \in N$ and y' in the S_N^y-block we automatically have $x' \prec_{\mathrm{B}} y'$. So at this stage of the construction we need not be concerned about the criticality w.r.t. lower order nodes. The procedure of "making an S_N^y-block" is now described in detail.

- Find x_0 such that $x_0 \in N$ and $x_0 R_e^B y$ for some B. This x_0 is unique if it exists and will be fixed in the sequel.

- At the start the S^y_N-block is nothing but the empty set.
- Provide every x' in $N \setminus \{x_0\}$ for which there is some deficiency of x' with respect to y, with a y' that resolves that particular deficiency in a way that $y \subset_\Box y'$. This can be done seeing that $x \prec_\bot y$ and applying lemma 4.6 for a $\bot$-critical successor. Again y' is chosen to be maximal w.r.t. $\Box$-inclusion. Include this y' in the S^y_N-block and define moreover $yS_{x'}y'$.
- Close off under the *ILM* frame conditions in a minimal way.

Repeat the following steps as long as deficiencies wherever in N with respect to some element of the S^x_N-block exist.

- Take a pair $x' \in N$ and y' in the S^y_N-block such that $x'Ry'$ and there exists some deficiency of x' w.r.t. y'. Say $C \rhd D$ is this deficiency and $C \in y'$. First look if there is some world y'' in the S^y_N-block such that $D \in y''$ and both $x'Ry''$ and $\forall E(\Box E \in y' \rightarrow \Box E \in y'')$. If this is so, define $y'S_{x'}y''$ and the deficiency has disappeared. If it is not possible to find such a y'', it must be created and added to the S^y_N-block. (Note again that checking whether a relevant y'' exists already is only necessary for establishing the finite model property; if we are just after completeness we can skip this part.) Using lemma 4.8 and the stipulation that $x' \prec_\bot y'$ guarantees the existence of such an object. Again we define in this case $y'S_{x'}y''$.
- Close off under the *ILM* frame conditions.

The same observations as before insure the termination of this procedure.

4.3.4 Correctness and termination

Troughout the whole contruction invariants played an important role. It has not yet been really proved that they indeed are invariants. As the construction method is inductively defined, we will prove the invariants by induction on the construction of the model. So, proving the correctness of an invariant consists of showing that it holds at the basis, i.e. the model $\langle\{m_0\}, \varnothing, \varnothing, \Vdash\rangle$, and show that it is preserved in the inductive steps. Two inductive steps will be considered: eliminating a problem and eliminating a deficiency. Eliminating a deficiency in itself is an inductive process but it will not be necessary to consider it in total detail while proving the correctness of an invariant.

The proofs of the correctness of the invariants are quite laborious and do not involve any profound mathematics. We add them though, in order to provide a complete proof. The already convinced reader can just skip this part or skim it over.

The first invariant we met was $(*) : xSy \to x \subset_\Box y$. To prove this invariant we take the conjunction of this statement with **invariant 2**: $xRy \to x \prec y$. The inductive proof of this conjunction will run as follows.

- In the basis model there are no R or S relations, so both conjuncts are automatically satisfied.
- Suppose a new y is defined when eliminating a problem, say $xR_e^B y$, and the model is closed under the frame conditions in a minimal way. Note that $x \subsetneqq_\Box y$ and $\Box\neg B \in y \setminus x$. If the frame conditions demand $x'Sy$, this can only be because the frame conditions impose $x'Ry$ because of for example $x'Sx$. But in this case the induction hypothesis tells us $x' \subset_\Box x$, and as $x \subsetneqq_\Box y$, it is clear that $x' \subset_\Box y$. So $x'Sy \to x' \subset_\Box y$ is thereby established. As $x' \subset_\Box x$ and $x \prec y$, we have $\Box A \in x \to A, \Box A \in y$. In other words we have $x \prec y$ as well as $x' \prec y$.
- Suppose now that a new world y' is created while eliminating a deficiency in x w.r.t. y. The world is chosen so that $(*)$ holds. As yS_xy', one is obliged to define xRy'. But as y' is chosen by lemma 4.8 one has $x \prec y'$. It is easy to see that if $x'Ry'$ is imposed by the frame conditions then either $x'Sx$ or $x'Rx$ $(x' \neq x)$. If $x'Rx$ then $x' \prec x$ and by the transitivity of $\prec$, $x' \prec y'$. If $x'Sx$ then $x' \subset_\Box x$ and because of $x \prec y'$ one has $\Box A \in x' \Rightarrow \Box A \in x \Rightarrow A, \Box A \in y'$. Also there is a $\Box A \in y' \setminus x'$. So indeed $x' \prec y'$.

Invariant 1 is evidently seen to be true as all the time the finite construction is closed off under the frame conditions. And the frame conditions together with the construction method form a consistent process, i.e. never will one frame condition exclude the other.

Now we come to prove the fact that a B-critical cone of x lies indeed B-critically above x. For this is what **invariant 3** actually says: $y \in C_x^B \to x \prec_{\mathrm{B}} y$.

- Again the basis step is trivially fulfilled.
- Now consider the case that the frame conditions demand $y \in C_x^B$ for some x and B: y here is the newly defined world, which was introduced to eliminate a problem in x', i.e. for some B', $x' R_e^{B'} y$. Of course we may assume that $x \neq x'$, otherwise things are trivial. We distinguish two different situations:
  - $x' \in C_x^B$. This implies $x \prec_{\mathrm{B}} x'$ and combined with $x' \prec y$ this yields $x \prec_{\mathrm{B}} y$.
  - $x' \notin C_x^B$. That the frame conditions impose $y \in C_x^B$, must be because $x'' S x'$ for some $x'' \in C_x^B$. Hence we have $x'' \subset_{\Box} x'$. If now $A \rhd B \in x$ then $\neg A, \Box\neg A \in x''$ and so $\Box\neg A \in x'$ (because of $x'' S x'$) and thus $\neg A, \Box\neg A \in y$ $(x' \prec y)$. Thus $x \prec_{\mathrm{B}} y$. (It is also possible that $x = x'$ and $B = B'$ so that one trivially obtains $x \prec_{\mathrm{B}} y$.)
- Suppose now y is introduced while eliminating a deficiency, and $y \in C_x^B$. We again distinguish two situations:
  - $\exists x' \in C_x^B\ x' R y$. Then the induction hypothesis applies to x' so $x \prec_{\mathrm{B}} x'$. But because $x' \prec y$, the required $x \prec_{\mathrm{B}} y$ is herewith confirmed.
  - $\neg(\exists x' \in C_x^B\ x' R y)$. In this situation one must have $y_0 S_x y$ for some $y_0 \in C_x^B$. By scrutinizing the process by which a deficiency is eliminated one must conclude that at some stage y has been added to the model by applying lemma 4.8 in making an $S_x^{y'}$-block. This means that y indeed lies B-critically above x.

Instead of proving **invariant 4**: $B \neq B' \to C_x^B \cap C_x^{B'} = \varnothing$, we will prove something stronger. First we define a new set $\tilde{C}_x^B$, of successors of x. We take $\tilde{C}_x^B$ the smallest set such that: $x R_e^B y \to y \in \tilde{C}_x^B$ and $y \in \tilde{C}_x^B \wedge y S y' \to y' \in \tilde{C}_x^B$. It is quite easy to show with induction on this definition that the invariant actually holds for $\tilde{C}_x^B$. That is $B \neq B' \to \tilde{C}_x^B \cap \tilde{C}_x^{B'} = \varnothing$.

- The basis model again is trivial.
- Suppose that y is added to the model while eliminating a problem in x' and the frame conditions force y to be in $\tilde{C}_x^B$. If $x \neq x'$ this can only be the case if for some $x'' \in \tilde{C}_x^B$ one has $x'' S x'$ and $x'' R y$ is thus forced. This also implies that $x' \in \tilde{C}_x^B$. By the same means one

concludes $x' \in \tilde{C}_x^{B'}$ if $y \in \tilde{C}_x^{B'}$. The induction hypothesis then gives $B = B'$. If $x = x'$ and the frame conditions impose $y \in \tilde{C}_x^B$ then this can only be because of $xR_e^B y$. It is clear then that y can't be in $\tilde{C}_x^{B'}$ for $B' \neq B$.

- Consider a new world y' which is added to eliminate a deficiency in x' w.r.t. y and suppose that the frame conditions demand $y' \in \tilde{C}_x^B$. Close inspection of the process of eliminating a deficiency yields the conclusion that y already must have been in $\tilde{C}_x^B$. The induction hypothesis tells us that y can not be in $\tilde{C}_x^{B'}$ for another B'. So the same holds for y'.

Instead of **invariant 5**: $ord(x) = ord(y) \wedge x \neq y \to C_x^B \cap C_y^{B'} = \varnothing$ we will prove the stronger invariant obtained by replacing C_x^B by $\tilde{C}_x^B$ in 5.

- The invariant is obviously satisfied in the basic model.
- Suppose that y is introduced to eliminate a problem in x'' and suppose moreover that the frame conditions demand $y \in \tilde{C}_x^B$. If $x'' \in \tilde{C}_x^B$ then the induction hypothesis tells us that $x'' \notin \tilde{C}_{x'}^{B'}$ for any $x' \neq x$ and $ord(x') = ord(x)$. So the same holds for y. If $x'' \notin \tilde{C}_x^B$ and yet $y \in \tilde{C}_x^B$ is obligatory, it must be that $x'' = x$ and $xR_e^B y$. $ord(x') = ord(x)$ implies $x \notin \tilde{C}_{x'}^{B'}$ for any B'. So $y \in \tilde{C}_{x'}^{B'}$ for $x \neq x'$ is not possible.
- The case when a deficiency is introduced is completely analogous to the previously treated invariant 4.

Overall correctness and termination By proving the correctness of the invariants we have also justified the claims based on them. It remains to show that the construction is overall correct. Indeed problems and deficiencies are eliminated successively, but is this sufficient? Once a deficiency in x w.r.t. y is eliminated, it is clear that it will never turn up again. Every new stage of the model is properly extending the previous one. So yS_xy', once there, will never disappear. A problem $\neg(A \rhd B)$ in x is eliminated by creating a B-critical successor y where A holds. Invariant 3 implies that $\forall y'(yS_xy' \to x \prec_{\mathrm{B}} y')$. This means that at every new stage of the model no problems that have already been dealt with will reoccur.

The construction can be seen as repeatedly eliminating a problem and then all the new deficiencies that emerge by tackling the problem. It should be noted that, during all of this process, the new worlds that are added are of higher order then the order of the world in which the problem occurred. This gives an inductive flavor. And indeed one can show that the order has an upper bound. For if $ord(x) = |BR(A)|$, one has $\Box\neg A \in BR(A) \to \Box\neg A \in x$. But as $\Box\neg A \vdash A \rhd B$, also $A \rhd B \in x$. This also holds for all relevant formulas, so no problems will exist once $ord(x) = |BR(A)|$. In the model every non-circular path (using R and S transitions) will finally end up in such a top node. So the processs will indeed stop. (In other words, the strictly monotonic increase of the order function together with the upper bound on the function enforces the termination of the process.)

5 The modal completeness of *ILM₀* via the construction method

5.1 General outline of the proof

By the same means by which the modal completeness of *ILM* was verified, we now obtain a new result: the modal completeness of ILM_0. Again a countermodel is built by successively eliminating problems and deficiencies. Since in this case we have not succeeded in proving the finite model property we will not strive for termination of the construction. In the new case we have to end up with an ILM_0-frame. Let us recall the M_0-axiom and its corresponding frame condition. M_0 is the axiom $A \rhd B \to \Diamond A \wedge \Box C \rhd B \wedge \Box C$ and its frame condition is:

$$vRwRxS_vyRz \to wRz$$

In order to obtain this property in our model consisting of copies of maximal ILM_0-consistent sets of modal sentences, it suffices to choose y such that $w \subset_\Box y$. It turns out to be possible to do so. Fix again an $\mathcal{A}$ such that $ILM_0 \nvdash \mathcal{A}$. (Again this font is a little unusual in the modal setting, but we again want to have the A free to use in the course of this chapter.) The set of relevant formulas R is as before, the smallest set of formulas containing $\mathcal{A}$ and closed under taking subformulas and single negations. As $\mathcal{A}$ is not derivable in ILM_0 there exists a maximal ILM_0-consistent set m_0 containing $\neg\mathcal{A}$. A model $\mathcal{M}$ is built above this m_0 so that the truth lemma will hold for all relevant formulas, and thus $\mathcal{M}, m_0 \Vdash \neg\mathcal{A}$. We have to provide ourselves with some tools before we can start with the actual construction.

5.2 Tools

Definition 5.1 $\widetilde{M_0} := \{\Gamma \mid \Gamma$ *maximal* ILM_0*-consistent set of modal sentences s.t.* $\forall \mathrm{A} \in \Gamma \;\; Prop(\mathrm{A}) \subset Prop(\mathcal{A})\}$. $Prop(B)$ *is the set of propositional variables occurring in* B.

The definitions of $\prec$ and $\Gamma \prec_{\mathrm{B}} \Delta$ are precisely as before.
In perfect analogy with the case of *ILM* we have the following two lemmas:

Lemma 5.2 *Let* $x \in \widetilde{M_0}$ *and* $\neg(A \rhd B) \in x \cap R$. *There exists* y *such that* $x \prec_B y$ *and* $A \in y$. *Moreover,* y *can be chosen to contain a "maximal amount" of* $\Box$*-formulas.*

Lemma 5.3 *Let* $x \in \widetilde{M_0}$ *with* $A \rhd B \in x$ *and let* $y \in \widetilde{M_0}$ *be such that* $x \prec_C y$ *and* $A \in y$. *There exists* $z \;{}_C{\succ}\; x$, *with* $B \in z$.

The M_0-axiom is essentially used only in the next lemma.

Lemma 5.4 *Consider* $w \prec_B x \prec y$, *all in* $\widetilde{M_0}$, *such that* $C \rhd D \in w$, *and* $C \in y$. *Then there exists* $z \;{}_B{\succ}\; x$ *with both* $D \in z$ *and* $x \subset_{\Box} z$. *This* z *can be chosen to be maximal w.r.t. the* $\Box$*-inclusion.*

Proof of Lemma 5.4. Consider the following set:

$$X := \{A, \Box A \mid \Box A \in w\} \cup \{\neg E, \Box\neg E \mid E \rhd B \in w\} \cup \{\Box F \mid \Box F \in x\} \cup \{D\}.$$

We will prove that X is consistent. For suppose it were not, then by completeness we have for a certain finite subset of X that

$$\begin{array}{ll} A_1, \ldots, A_l, \Box A_1, \ldots, \Box A_l, \Box F_1, \ldots, \Box F_m, & \\ \neg E_1, \ldots, \neg E_n, \Box\neg E_1, \ldots, \Box\neg E_n, D & \vdash \;\; \bot \end{array} \tag{2}$$

As $C \rhd D \in w$ and $w \in \widetilde{M_0}$, we also have $\Diamond C \wedge \bigwedge_{i=1}^{m} \Box F_i \rhd D \wedge \bigwedge_{i=1}^{m} \Box F_i \in w$. Because $x \prec y$ and $C \in y$, one also has $\Diamond C \in x$ and hence also $\Diamond C \wedge \bigwedge_{i=1}^{m} \Box F_i \in x$. Now applying lemma 5.3 yields a $z \in \widetilde{M_0}$ with $D \wedge \bigwedge_{i=1}^{m} \Box F_i \in z$ while $w \prec_{\mathrm{B}} z$, but this certainly conflicts with (2). As before one can see that the z can be chosen to be maximal w.r.t. $\Box$-inclusion. QED

5.3 The construction

The main body of the construction procedure for obtaining our required $\mathit{ILM_0}$-model is quite analogous to the case of ILM. One difference is that we now deal with copies of maximal $\mathit{ILM_0}$-consistent sets, and that we have to satisfy another frame condition. The nomenclature will thus be exactly the same. In the case of $\mathit{ILM_0}$ we will not have "there are no deficiencies" as an invariant. We start again with an $\mathit{ILM_0}$-consistent set m_0 containing $\neg\mathcal{A}$. If $\mathit{ILM_0} \nvdash \mathcal{A}$. Again the only entity that can be defined globally is the forcing relation $\Vdash$. We set $x \Vdash P \Leftrightarrow P \in x$, for the propositional variables. The body of the procedure will now be as follows:

- As a first approximation of the required model, set its domain to be $\{m_0\}$ and $R = S = \varnothing$.

begin
As long as problems or deficiencies still do exist in the model, enter this loop:

- If a problem does exist, then pick one of lowest order and eliminate this very problem. After having done so, close off under the frame conditions.
- If some deficiency exist somewhere, then fix an x and a y, such that deficiencies in x w.r.t. y do exist. Eliminate all these deficiencies and close off under the frame conditions.

end

We choose the way of eliminating problems, respectively deficiencies, cleverly so as to have some useful properties in the model. The useful properties that we need in our model are stated again as invariants as they hold at the beginning and at the end of each loop.

1. The frame is an $\mathit{ILM_0}$-frame.
2. $xRy \to x \prec y$.
3. $xRx'RyS_xy' \to x' \subset_\Box y'$.
4. $y \in C_x^B \to x \prec_\mathrm{B} y$.
5. $B \neq B' \to C_x^B \cap C_x^{B'} = \varnothing$.

6. $xRyRz \to [\exists\, y'\ xRy'Rz \land \forall y''(xRy''Rz \to y'' \subset_{\Box} y')]$.

Again one should run through the whole construction checking that these are indeed invariants. Surely they hold at the beginning of the loop. So we have to make sure that after every execution of the loop all the invariants hold.

5.3.1 Problems

A problem is dealt with just as before. If the problem in x is $\neg(A \rhd B)$ you find a world y applying lemma 5.2 such that $x \prec_{\mathrm{B}} y$, $A \in y$ and y is such that it is also maximal w.r.t. the $\Box$-inclusion. You then define $xR_e^B y$ and close off in a minimal way under the ILM_0-frame conditions.

5.3.2 Deficiencies

Again the deficiencies are harder to deal with. A **first difficulty** you encounter is in incorporating the ILM_0-frame condition. So, if you eliminate a deficiency in x w.r.t. y by defining yS_xy', you should make sure that every possible successor of y' can also be a successor of some intermediate x' (intermediate means $xRx'Ry$). This is done, as stated before, by demanding $x' \subset_{\Box} y'$. Lemma 5.4 tells us that this is indeed possible. So for every y'' with $y' \prec y''$, you immediately have $x' \prec y''$. A closer examination brings us to the **second difficulty**, that is, that there might be more intermediate worlds between x and y. But the invariant 6 tells us that there is always an intermediate world that is maximal w.r.t. the $\Box$-inclusion relation. This settles the second difficulty if we just apply lemma 5.4 every time to the $\Box$-maximal intermediate world. A **third difficulty** is found when realizing that the S-relation must be transitive. Say $C \rhd D$ is a deficiency in x w.r.t. y and we have $x \prec_B x'Ry$. We want to create a B-critical successor y' of x with $D \in y'$ and $x' \subset_{\Box} y'$. By doing so, every successor of y' can automatically be defined a successor of x' and the ILM_0-frame condition is satisfied. Invariant 6 tells us that without loss of generality we can take this x' to be maximal w.r.t. the $\Box$-inclusion relation. Creating y' containing D, to eliminate the deficiency $C \rhd D$ in x w.r.t. y is done in such a way that $x' \subset_{\Box} y'$. The existence of this required entity is guaranteed by lemma 5.4. Automatically we now have $xRx''Ry \to x'' \subset_{\Box} y'$. We accordingly set yS_xy'. The general frame conditions demand xRy'. Now it may happen that $C' \rhd D'$ is a deficiency in x w.r.t. y'. It is easy to find a world y'' with D' in it and being B-critical above x. The transitivity of S_x forces us to also have yS_xy'', but by no means we can ensure that $x' \subset_{\Box} y''$. Somehow we have to relate the possible deficiencies to each other. As we have only finitely many relevant sentences of the form $C \rhd D$, the following lemma helps us out.

Lemma 5.5

$$
\begin{array}{llr}
E_0 \rhd F_0, \ldots, E_n \rhd F_n \vdash_{ILM_0} \Diamond E_0 \wedge \Box C \rhd & (F_0 \wedge \neg E_1 \wedge \ldots \neg E_n \wedge \Box C) & \vee \\
& \vee (F_1 \wedge \neg E_2 \wedge \ldots \neg E_n \wedge \Box C) & \vee \\
& \vdots \qquad\qquad \vdots & \vdots \\
& \vee (F_n \wedge \Box C). &
\end{array}
$$

Proof of Lemma 5.5. By simple propositional logic we have $\vdash E_0 \rhd F_0 \rightarrow E_0 \rhd (F_0 \wedge \neg E_1 \wedge \ldots \wedge \neg E_n) \vee (F_0 \wedge E_1) \vee \ldots \vee (F_0 \wedge E_n)$. As each $E_i \rhd F_i$ is assumed, we obtain $E_0 \rhd (F_0 \wedge \neg E_1 \wedge \ldots \wedge \neg E_n) \vee F_1 \vee \ldots \vee F_n$. Again we split up F_1 into $(F_1 \wedge \neg E_2 \wedge \ldots \wedge \neg E_n) \vee (F_1 \wedge E_2) \vee (F_1 \wedge E_3) \vee \ldots \vee (F_1 \wedge E_n)$. Then we note that $F_i \wedge E_j \rhd F_j$ so we obtain $E_0 \rhd (F_0 \wedge \neg E_1 \wedge \ldots \wedge \neg E_n) \vee (F_1 \wedge \neg E_2 \wedge \ldots \wedge \neg E_n) \vee F_2 \vee \ldots \vee F_n$. Proceeding like this yields $E_0 \rhd (F_0 \wedge \neg E_1 \wedge \ldots \wedge \neg E_n) \vee (F_1 \wedge \neg E_2 \wedge \ldots \wedge \neg E_n) \vee \ldots \vee F_n$. Now we apply the M_0 axiom to obtain the required result: $\Diamond E_0 \wedge \Box C \rhd (F_0 \wedge \neg E_1 \ldots \neg E_n \wedge \Box C) \vee \ldots \vee (F_n \wedge \Box C)$.
QED

As the difference between the various E_i's is not essential, one has the lemma for any permutation of the indices. If one encounters a deficiency in x w.r.t. y, say $E_0 \rhd F_0$, one proceeds as follows. First a list is made of all relevant sentences of the form $C \rhd D$. Let this list be $E_0 \rhd F_0, \ldots, E_n \rhd F_n$. You can consider these as the *possible deficiencies*. Let x' be such that $xRx'Ry$ and $xRx''Ry \rightarrow x'' \subset_\Box x'$. All deficiencies in x w.r.t. y will be eliminated in such a way that no new deficiencies will occur in x w.r.t. the newly created worlds. Again this process shall henceforth be referred to as the process of making an S^y_x-block. The $\Box$-inclusion of x' shall be taken into account while creating this S^y_x-block.
As stated before lemma 5.5 actually represents a manifold of statements. All of them will be used in eliminating the deficiency $E_0 \rhd F_0$. Let π be a permutation of $\{1, 2, \ldots, n\}$. With any π a series of sets Δ^π_i is defined. ($i \in \{0, 1, \ldots, n\}$.)

$$\Delta^\pi_0 = \{E_1, \ldots, E_n\}, \quad \Delta^\pi_{i+1} = \Delta^\pi_i \setminus \{E_{\pi(i+1)}\}.$$

For the sake of convenience we define $\forall \pi : \pi(0) = 0$. The notation $\neg \Delta^\pi_i$ stands for the sentence $\bigwedge_{E_j \in \Delta^\pi_i} \neg E_j$. With this terminology one can easily write down the $n!$ useful variants of lemma 5.5.

$$P_\pi := \Diamond E_0 \wedge \Box C \rhd \bigvee\!\!\!\bigvee_{i=0}^{n}(F_{\pi(i)} \wedge \neg\Delta_i^\pi \wedge \Box C).$$

For any P_π lemma 5.3 can be applied. Recall our situation, $xR_e^B x'Ry$ with $E_0 \rhd F_0, \ldots, E_n \rhd F_n$ the possible deficiencies in x w.r.t. y. So applying lemma 5.3 to any P_π gives a maximal *ILM$_0$*-consistent set containing one of the disjuncts of the consequent of P_π. Note that $P_\pi \in x$ because $x \vdash P_\pi$. We now form a set of disjuncts $\mathcal{D}$ as follows. For every P_π you choose the leftmost disjunct of the consequent which is realizable by applying lemmas 5.3 and 5.5.

Lemma 5.6 *Consider the above situation. Let $F_{i_1}, \ldots, F_{i_k}$ be all the formulas that do not occur in any of the $n!$ disjuncts in $\mathcal{D}$, and let F_j be an arbitrary formula of the possible F's which does occur in some disjunct in $\mathcal{D}$. There must be some disjunct in $\mathcal{D}$ where both F_j and $\neg E_{i_1}, \ldots, \neg E_{i_k}$ hold.*

Proof of Lemma 5.6. In order to show this, let P_{π^0} be a sentence from which a disjunct containing F_j was chosen for $\mathcal{D}$. If all the F_{i_l}'s occur in the disjunct on the right hand side of F_j, one also has $E_{i_l} \in \Delta_m^{\pi^0}$ for $l = 1, \ldots, k$. Here m is such that $\pi^0(m) = j$. Thus the disjunct under consideration already contains F_j and all of the $\neg E_{i_l}$'s.
Now suppose there is some F_{i_l} in a disjunct of the consequent (or we just say disjunct) of P_{π^0}, left from the disjunct where F_j occurs. We claim that interchanging the disjunct where the F_{i_l} occurs, with its direct neighbour to the right, does not essentially change the sentence w.r.t. $\mathcal{D}$. Two sentences are said to not essentially differ w.r.t. $\mathcal{D}$ if the disjuncts they add to $\mathcal{D}$ contain the same F_i. By interchanging two disjuncts we mean here a slightly different process thaen usual. $P_{\pi'}$ is the sentence obtained by interchanging the i-th and the $i+1$-th disjunct in the consequent of P_π if π' satisfies the following conditions:

$$\pi'(a) = \begin{cases} \pi(a+1) & \text{if } a = i \\ \pi(a-1) & \text{if } a = i+1 \\ \pi(a) & \text{otherwise} \end{cases}$$

It is quite obvious that interchanging the disjunct containing the F_{i_l} and its direct right neighbor does not essentially matter w.r.t. $\mathcal{D}$. Say $\pi^0(i) = i_l$. Now all the disjuncts remain the same except for the i-th and the $i+1$-th after interchanging. But clearly *ILM$_0$* $\vdash F_{\pi'(i)} \wedge \neg\Delta_i^{\pi'} \to F_{\pi^0(i+1)} \wedge \neg\Delta_{i+1}^{\pi^0}$.

So if $\pi^0(i+1) \neq j$, $F_{\pi'(i)} \wedge \neg\Delta_i^{\pi'}$ will certainly not be realizable. (Recall that this F_{i_l} appeared on the left-hand side of F_j!) $F_{\pi'(i+1)} = F_{i_l}$ and F_{i_l} is not in $\mathcal{D}$, so does not occur realizably in $P_{\pi'}$. So, still the leftmost disjunct in $P_{\pi'}$ that is realizable is the one containing F_j. It is clear that by repeatedly interchanging disjuncts like this, all the F_{i_l}'s can be pushed to the right of the disjunct containing F_j. The sentence $P_{\tilde{\pi}}$ finally obtained will still add a disjunct containing F_j to $\mathcal{D}$. In $P_{\tilde{\pi}}$ all the F_{i_l}'s occur to the right of F_j. So the disjunct containing F_j also contains $\neg E_{i_1} \wedge \ldots \wedge \neg E_{i_k}$. So indeed for every F_j that occurs in $\mathcal{D}$ there is also a disjunct in $\mathcal{D}$ containing $F_j \wedge \neg E_{i_1} \ldots \wedge \neg E_{i_k}$. QED

Eliminating deficiencies: the complete picture. We now prove by induction on the number of possible deficiencies that every deficiency in x w.r.t. y can be eliminated in an adequate way. (For the time being we only consider the harder case when intermediate worlds do exist.) Adequate means that all the S_x successors of y that are defined for this purpose contain all the $\Box$-formulas of x' $(xR_e^B x'Ry)$. The new worlds should not provoke new deficiencies in x. And all of them should lie B-critically above x.

- If you only have $E_0 \rhd F_0$ as possible deficiency in x and this is indeed a deficiency, i.e. $E_0 \in y$, you can apply lemma 5.4. This yields a y' such that $F_0 \in y'$, $x \prec_{\mathrm{B}} y'$ and $x' \subset_{\Box} y'$. This y is taken to be maximal w.r.t. the $\Box$-inclusion. You can thus safely define yS_xy' and no new deficiencies will emerge in x w.r.t. y, for they have to be amongst the possible deficiencies. The model is closed under the frame conditions.

- If the possible deficiencies are $E_0 \rhd F_0, \ldots, E_n \rhd F_n$, one can make use of lemma 5.5 to derive $n!$ useful sentences. Also the corresponding set $\mathcal{D}$ is formed by the same method as before. Let again $F_{i_1}, \ldots, F_{i_k}$ be all the F's that do not occur as such in $\mathcal{D}$. Lemma 5.6 guarantees the existence of a subset $\mathcal{D}'$ which has, for every F_j that occurs in $\mathcal{D}$, one disjunct that contains both F_j and $\neg E_{i_1}, \ldots, \neg E_{i_k}$. For every disjunct in $\mathcal{D}'$ a corresponding world can be found where that very disjunct holds. This world will also lie B-critically above x.
  But one does not meet all the requirements like this. For the worlds only contain a finite part of the box formulas in x', namely the $\Box C$ that was included in all the $n!$ formulas. This calls for compactness again. Let $\Box C_1, \Box C_2, \Box C_3, \ldots$ be an enumeration of all the box formulas in x'. Let $\Box X_i := \Box \bigwedge_{j=1}^{i} C_j$. For every $\Box X_i$ a set $\mathcal{D}'$ as above is formed. As only a finite number of partitions is possible for $\mathcal{D}'$, one specific

partition must show up infinitely many times. With respect to every individual world of this (or better corresponding to this) partition the compactness theorem can be applied to obtain a set of worlds. All of these worlds will contain some F_j, all of $\neg F_{i_1}, \ldots, \neg F_{i_k}$, and all the $\Box C \in x'$. They can also be taken maximal w.r.t. $\Box$-inclusion. This set of worlds is called $\mathcal{D}^y_{x,x'}$. So $F_{i_1}, \ldots, F_{i_k}$ do not occur in $\mathcal{D}^y_{x,x'}$. For these remaining k possible deficiencies you can apply the induction hypothesis while ignoring the $n-k+1$ other possible deficiencies. Note that if $k = n+1$ the whole situation becomes very trivial. So, without loss of generality, we can assume that $k < n+1$. Applying the induction hypothesis yields a whole net of B-critical successors of x solving all the k deficiencies disregarding the other $n-k+1$ possible deficiencies. Every world in this net contains all the $\Box$-formulas of x'. All the individual members of this net are maximal w.r.t. $\Box$-inclusion. By "attaching" this B-critical net to y, k deficiencies are solved but some deficiencies among the other $n-k+1$ possible ones may "have become active". You eliminate these deficiencies by defining an S_x arrow from every world in the B-critical net to every world in $\mathcal{D}^y_{x,x'}$. S_x on $\mathcal{D}^y_{x,x'}$ is defined such that $\mathcal{D}^y_{x,x'}$ is completely S_x-connected. After having done so, no deficiency in x remains w.r.t. any of the newly defined worlds. If $E_i \rhd F_i$ belongs to the group of k it is, if necessary, eliminated in the B-critical net. If $E_i \rhd F_i$ belongs to the group of $n-k+1$, there is always an S_x-arrow to a world in $\mathcal{D}^y_{x,x'}$ where F_i holds. In this world E_j cannot hold for any $E_j \rhd F_j$ from the group of k. So indeed all the deficiencies are eliminated. Moreover transitivity of the S_x relation is guaranteed. The process described here will be referred to as the process of making an S^y_x-block.

If there is a deficiency in x w.r.t. y and *there is no intermediate world*, the situation becomes even easier. Again the process of eliminating this deficiency in x w.r.t. y shall be referred to as the process of making an S^y_x-block. If in a general setting the process of making an S^y_x-block is mentioned, either one of these two is meant, depending on the presence of intermediate worlds. The process of making an S^y_x-block in the case of no intermediate worlds can inductively be described as follows.

- Let y be in the S^y_x-block.
- **begin** As long as deficiencies in x exist w.r.t. some world y in the S^y_x-block, apply lemma 5.2 to obtain y'' such that y'' resolves this

deficiency by setting $y'S_xy''$ and xRy''. As the whole S_x^y-block is in at most one B-critical cone of x, there is no ambiguity in applying lemma 5.2. Now make the whole S_x^y-block completely S_x-connected.
end

As there is only a finite number of possible deficiencies, this process must come to an end.

The actual construction. All the above ingredients make sure that the logic is complete. However there is no ingredient that would yield the finiteness of the model. In proving the completeness of *ILM₀* an infinite model is thus provided as the limit of an iterative process. In the limit neither problems nor deficiencies should be present. Most likely, eliminating a problem or a deficiency will create many new problems and deficiencies as a side effect. It should be made sure that every problem or deficiency is at some stage eliminated. This can be done by adequate labeling as is also done in [MVar]. By means of a set I, the set of *imperfections* we will keep track of all the problems and deficiencies which have not yet been eliminated. The model construction can thus be represented by the following procedure.

- The first approach to the model will be to take the domain $\{m_0\}$ and $R = S = \varnothing$. All the problems of m_0 are stored in I.

As long as I is nonempty **repeat** the following actions.

- Select the oldest member of I. If this is not uniquely defined, just pick arbitrarily one of the oldest. Old refers of course to how many repetitions the element has already been in I.
  - If this oldest member of I is a problem $\neg(A \rhd B)$ in some world x, eliminate it by defining xR_e^By for an adequate y. This y is provided by lemma 5.2 and as usual is chosen such that $x \prec_{\mathrm{B}} y$ and $A \in y$.
  - If the oldest member of I is a deficiency in some world x w.r.t. some world y, then eliminate it by making an S_x^y-block.
- Close off under the *ILM₀* frame conditions in a minimal way. Add the new freshly born problems and deficiencies to I.

This construction method produces a whole series of *ILM$_0$*-models $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \mathcal{M}_3 \subset \ldots$. With each execution of the repeat loop the previous model is extended. In general, for no $n \in \omega$, $\mathcal{M}_n = \mathcal{M}_{n+1}$ will hold. But one can consider $\mathcal{M}_\infty := \bigcup_{i\in\omega} \mathcal{M}_i$. It is clear that in $\mathcal{M}_\infty$ neither problems nor deficiencies hold. This of course under the assumption that all the invariants we used throughout the construction are indeed invariants, and also hold for the infinite model. If this is true, then problems that have been eliminated will never re-emerge. This is reflected by invariant 4. The main strategy for proving the invariants will be to prove their correctness for all $\mathcal{M}_n$, and then show that this extends to the infinite model. Before doing so it is useful to note some features of the chain of models $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \mathcal{M}_3 \subset \ldots$. First of all one can say there is a uniform upper bound to the height of the model. (The height is defined as the maximal length of some chain $x_0Rx_1R\ldots Rx_n$ in the model.) Therefore $\mathcal{M}_\infty$ also has finite height. Thus $\mathcal{M}_\infty$ is conversely well-founded when it comes to the R-relation. Second and also important, one can see that $\mathcal{M}_{n+1}$ is sort of an "end extension" of $\mathcal{M}_n$ in the sense that no intermediate worlds will be added. So, if $xTy \in \mathcal{M}_n$ and $z \in \mathcal{M}_{n+1} \backslash \mathcal{M}_n$, then for no z, $xTzTy$ for both $T = S_{x_i}$ as well as for $T = R$. Further it is clear that all defined notions like R, S, C_x^B, etc. are weakly monotonic increasing entities.

Correctness. It is quite clear that the model $\mathcal{M}_\infty$ is an *ILM$_0$*-model. To be an *ILM$_0$*-model, the frame induced by $\mathcal{M}_\infty$ must be an *ILM$_0$*-frame, i.e. it must be an IL-frame and it must satisfy the typical *ILM$_0$* frame condition. The model $\mathcal{M}_\infty$ has already seen to be conversely well-founded w.r.t. the R-relation. Actually the well-foundedness is the only frame condition which is not expressible by a first order quantor rank four formula. (Of course it is not at all f.o. expressible.) So if one of the other frame conditions were not true, some witnesses could be found. Now suppose there were w, x, y and z which falsify one of the frame conditions. Then already for some n_0 one has $w, x, y, z \in \mathcal{M}_{n_0}$. But just as before, it is easy to see that every $\mathcal{M}_n$ defines an *ILM$_0$*-frame hence also $\mathcal{M}_{n_0}$. This conflicts the assertion that $\mathcal{M}_\infty$ would not be an *ILM$_0$*-model. So indeed $\mathcal{M}_\infty$ defines an *ILM$_0$*-frame.

Invariants 2, 3 and 6 are proved inductively together. It is shown they hold in every $\mathcal{M}_n$.

- It is trivial as usual that all the invariants hold in the basis model.

- Suppose a new world y is added by eliminating a problem in x by setting xR_e^By.

  - It is clear that $x \prec y$. If $x'Ry$ then either $x'Rx$, $x' = x$ or x' is incomparable to x. In the first case it is easy to see that $x' \prec y$. In the latter case it must be so that $x'Rx''S_{x_0}x$ for some x', x'' and x_0. Note that this x_0 is uniquely defined. We can take x'' such that x was introduced to eliminate a deficiency in x_0 w.r.t. x'' while taking intermediate points into account. So $x' \subset_\Box x$. (Here we use the induction hypothesis for invariant 6.) But as xRy and thus $x \prec y$, also $x' \prec y$.
  - Invariant 3 is preserved while eliminating a problem, for the only new S relation that is added, is by the closure under the frame conditions: $x'Ry' \to x'Sy'$; but then clearly $x' \subset_\Box y'$.
  - Now consider invariant 6 after the new world y has been added to eliminate a problem in x. The only new and interesting case to consider will be if these worlds x and y are involved, for example when x_0Rx_1Ry for some x_0 and x_1. Close inspection of the situation results in concluding $x_1 \subset_\Box x$. If x_1Rx or $x_1 = x$ this is clear. If neither of these is true, it must be so that $x'Rx_1Rx''S_{x'}x$ for some x'' and unique x'. x'' can be chosen such that x has been introduced in the model to resolve a deficiency in x' w.r.t. x''. This has been done by the procedure of making an $S_{x'}^{x''}$-block. So the intermediate world with the highest amount of $\Box$ formulas has been used in this procedure (Induction!). Consequently $x_1 \subset_\Box x$. Indeed x is the world maximal w.r.t. $\Box$ inclusion.

- Now suppose a new world y' is added by eliminating a deficiency in x w.r.t. some world y.

  - If for some world x_0, x_0Ry' is demanded, it can only be the case that x_0Rx. Clearly $x \prec y'$ and, by induction $x_0 \prec x$, so $x_0 \prec y'$.
  - Consider the case when $xRx'RyS_xy'$ with x, y, and y' still standing for the same designated symbols. As y' was introduced by applying the procedure of making an S_x^y-block, we can be sure that the intermediate world x'' maximal w.r.t. $\Box$-inclusion has been used in applying this procedure (induction again), so that $x'' \subset_\Box y'$. Consequently $x' \subset y'$.
  - Invariant 6 is easily seen to be true. x_0Ry' can occur only when x_0Rx. So every "R path" ending up in y' must pass through x.

Consequently x is maximal w.r.t. the $\Box$-inclusion in this particular case.

This concludes the proof that invariants 2, 3 and 6 hold in every $\mathcal{M}_n$. It is easy to see that invariants 2 and 3 extend to the infinite model. Recalling our earlier consideration of "end extensions", forces us to conclude that invariant 6 extends to the infinite model as well.

It is quite easy to see that invariant 4 holds in every $\mathcal{M}_n$. The basic model trivially satisfies 4. If some y is added to the model to eliminate a problem in x', either x' is in C_x^B or it is not. Only the latter case needs some argument. If $y \in C_x^B$ and $x \notin C_x^B$, there must be some $x_1, x_2 \in C_x^B$ and some $x_0 R x$ such that $x_0 R x_1 R x_2$ and $x_2 S_{x_0} x'$. x_2 can be chosen so that x was introduced to eliminate a deficiency in x_0 w.r.t. x_2. Hence $x_1 \subset_\Box x'$. As $x_1 \subset_\Box x'$ and $x' \prec y$, one has $x \prec_{\mathrm{B}} y$. (Recall that $x \prec_{\mathrm{B}} x_1$.) The situation is very easy if a new world is added by eliminating a deficiency. Thus every $\mathcal{M}_n$ satisfies invariant 4. The notion of criticalness also extends to the infinite model. Analogously invariant 5 is seen to be true. (For example one could apply the same strategy as in the case of *ILM*.) The verification of the invariants concludes the completeness proof.

5.4 Some remarks on decidability

Our result on *ILM$_0$* does not include a decidabilty result. Some attempts have been made though. The finite model property would be sufficient for the decidability. In order to keep the model finite one should re-use worlds as was done in the completeness proof of *ILM*. Attention should be payed at the invariants. They must be preserved! Probably it will be necessary to also label the S_x-transitions throughout the construction.

6 A new principle

6.1 The birth of a new principle: P_0

When the research to the modal completeness of *ILM$_0$* was renewed by the author, it was suggested that *ILM$_0$* might be modally incomplete. Certainly this would not be the first modally incomplete principle. Albert Visser tried to strengthen the frame condition of *ILM$_0$* to arrive at a stronger principle. The frame condition of *ILM$_0$* is:

$$x_0 R x_1 R x_2 S_{x_0} y R y' \to x_1 R y'.$$

Instead of demanding an R-relation between x_1 and y', one can demand an S_{x_1}-relation between x_2 and y'. As we have $x_2 S_{x_1} y'$, we must also have $x_1 R y'$, so indeed the frame condition is hereby strengthened. The corresponding principle is readily found and baptized with the lyrical name of P_0.

$$P_0 \ : \quad A \rhd \Diamond B \to \Box(A \rhd B).$$

At first M_1 was suggested as a name, but at second thought P_0 seemed to be more appropriate. The reason is given below. As P_0 turns out to be an arithmetically valid principle one is obliged to subject it to a modal and comparative analysis. The target logic is the interpretability logic of all reasonable arithmetical theories, abbreviated *GIL*. As P_0 is a new generally valid principle, it brings us one step closer to *GIL*.

6.2 The enclosure of *GIL*

The new principle was finally named P_0 by Frank Veltman. He called it thus for its similarity with $P : A \rhd B \to \Box(A \rhd B)$. P_0 as he noticed can be seen as a weakening of P. Strengthening the antecedent of a conditional weakens it. And indeed $\vdash_{IL} A \rhd \Diamond B \to A \rhd B$. Similarly M can be weakened to obtain M_0. This calls for a general approach. On the one hand we have principles whose corresponding logic is too weak to be *GIL*, and on the other hand we have the logics *ILM* and *ILP* which are clearly too strong to be a candidate for *GIL*. The general approach would be to weaken the logics which are too strong and to strengthen the logics which are too weak. We can alter our logics in three different aspects:

- On an arithmetical basis one can appropriately vary the principles. This way of enclosing *GIL* is the most direct. It is quite difficult to think of candidate principles though and is outside the realm of this thesis.

- Another possibility can be to strengthen for example the frame-conditions of some principle which is too weak. *ILP₀* was found like this. For example one might try to strengthen the frame condition of W to demand $S \circ R$ conversely well-founded.

- The other possibility would be syntactical modulation. Many statements in interpretability logic take the form of a valid conditional.

Even the $\rhd$ operator can be considered as a conditional. Weakening a conditional can be done by weakening the consequent or strengthening the antecedent. For weakening or strengthening for example a consequent, one can use implications. The antecedent of a conditional is stronger than the consequent. All the axioms of *IL* can be written as an implicational statement and hence can be used to modify the other principles. $J5$ can be written in the equivalent form:

$$J5' : \quad A \rhd B \to \Diamond A \rhd B.$$

Using this form of $J5$ to modify other statements gives interesting results. M_0 can be obtained as a right weakening of M and P_0 can be seen as a left weakening of P. It has not yet been systematically investigated if new interesting principles can be found proceeding like this.

6.3 Independence results concerning P_0

In this paragraph the new principle will be compared to the previous principles from a modal viewpoint. Our modal semantics, Veltman models, does not provide a sufficiently strong tool for our analysis. As far as Veltman models are concerned P_0 implies M_0. But without modal completeness for *ILP₀* we do not know whether the notion of semantical consequence coincides with the notion of derivable consequence. In other words we can say nothing about the derivability of M_0 in ILP_0 on these grounds. And there was no reason for conjecturing $ILP_0 \vdash M_0$. On the contrary, the odds were against this inference. A nice comparison leads the intuition. *ILM* is known to be modally complete with respect to its class of characteristic frames. Another principle sometimes called KM_1, see [Vis90], has the same frame condition, but is not equivalent to M over *IL*. Here KM_1 is the formula $A \rhd \Diamond B \to \Box(A \to \Diamond B)$. Indeed, below ILP_0 will be shown to be incomplete.

IL is known to be sound w.r.t. Veltman semantics. That is, every derivable principle holds on all Veltman frames. Close inspection of the soundness proof shows the enormous amount of freedom one has in defining the semantics for *IL*. Various generalizations of Veltman semantics are known. (See e.g. [Šve91].) We will introduce here the notion of an IL_{set}-frame, an idea of Dick de Jongh for catching Svejdar's models in a general notion, as it was presented by Rineke Verbrugge in an unpublished document [Ver]. In the classical Veltman semantics there is the S relation. One can have $xS_{x_0}y$.

The y here is another world in the model. The main idea in the IL_{set} semantics is to replace this y with a set of worlds. So, we could have $xS_{x_0}\{y\}$ for example. One can now define $x \Vdash A \rhd B \Leftrightarrow \forall y(xRy \to \exists Y(yS_xY \wedge \forall y' \in Y(y' \vdash B)))$ and still have IL sound w.r.t. the new semantics. As we work with an existential quantifier in the old definition we will exclude the empty set as a possible S-successor: the axiom $A \rhd B \to (\Diamond A \to \Diamond B)$ demands $S_x \subset x\uparrow \times (\wp(x\uparrow) \setminus \{\varnothing\})$. Again the axiom $\Diamond A \rhd A$ demands that R can somehow be seen as embedded in S_x. In our case this reads $xRyRz \to yS_x\{z\}$. The axiom $\Box(A \to B) \to A \rhd B$ imposes a sort of reflexivity on our semantics; that is $yS_x\{y\}$. The transitivity clause leaves a lot of choice. The axiom states: $A \rhd B \wedge B \rhd C \to A \rhd C$. There is no first choice in how to adapt transitivity in the new semantics. It is sufficient to set $y_0S_xY \to \exists y' \in Y(\forall Y'(y'S_xY' \to y_0S_xY'))$. One could also replace the existential quantor by a universal quantor to obtain the definition of [Ver]. This will be our choice as well. Another possibility would be to demand $yS_xY \to \forall Z(\forall z(z \in Z \leftrightarrow \exists y' \in Y \exists Y'(yS_xY' \wedge y' \in Y \wedge z \in Y')) \to yS_xZ)$. The axiom $A \rhd C \wedge B \rhd C \to A \vee B \rhd C$ did not impose anything on the old semantics and the axiom maintains this special status. Thus we have:

Definition 6.1 *An IL_{set}-frame is a triple $(W, R, \{S_w \mid w \in W\})$ satisfying the following properties.*

- *(W, R) is an L-frame, that is, W is a nonempty set and $R \subset W \times W$ is such that R is transitive, irreflexive and conversely well-founded.*
- *For each w in W we have*
  - $S_w \subset w\uparrow \times (\wp(w\uparrow) \setminus \{\varnothing\})$
  - $wRx \to xS_w\{x\}$
  - $wRxRy \to xS_w\{y\}$
  - $y_0S_xY \to \forall y' \in Y\ \forall Y'(y'S_xY' \to y_0S_xY') \qquad (*)$

Definition 6.2 *An IL_{set}-model is a pair $(F, \Vdash)$ where F is an IL_{set}-frame and $\Vdash$ a forcing relation between worlds and proposition letters. The forcing relation is extended to sentences in the usual way when it comes to connectives or boxes. Furthermore the $\rhd$-operator is incorporated by $x \Vdash A \rhd B \Leftrightarrow \forall y(xRy \wedge y \Vdash A \to \exists Y(yS_xY \wedge \forall y' \in Y(y' \Vdash B)))$.*

This new semantics yields strong enough a tool to allocate the principle P_0 in the landscape of other interpretability principles. It turns out that P_0 has the highest possible degree of independence with respect to the principles M_0 and W. This result is stated in the next theorem.

Theorem 6.3 *The principles M_0, W and P_0 are maximally incomparable, that is to say it is not possible to derive one of the principles over IL using the remaining two as axiom schemes.*

PROOF OF THEOREM 6.3. We split the proof up in $\binom{3}{2} = 3$ different parts and make extensive use of countermodels. In depicting these models only the essential vertices will be shown. The relations imposed by the frame conditions (transitivity, reflexivity, etc.) will be omitted. The R-relation is represented by straight arrows, whereas the S_x-relation is represented by curly indexed arrows. Subsets of the universe are represented by encircling its members by a continuous closed curve. So by depicting a model we actually mean the smallest extension of the picture satisfying all the frame conditions. This is uniquely defined. Here the choice for the universal quantor in $(*)$ becomes clear. Of course we will use classical Veltman semantics if the principles characterize different properties on the frame. This is so in the first two cases.

- $ILM_0 W \nvdash P_0$.
  Suppose $ILM_0 W = ILW^*$ would prove P_0.

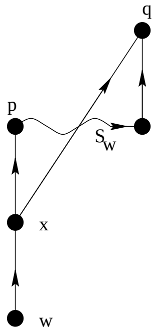


Figure 1. gives a countermodel using classical semantics to this implication. As it is an ILW^*-frame, $W^* = M_0 W$ is automatically fulfilled in w. We also have $w \Vdash p \rhd \Diamond q$. But clearly $w \nVdash \Box(p \rhd q)$ as $x \nVdash p \rhd q$.

Figure 1.

- $ILP_0 M_0 \nVdash W$.
  This is shown similarly.

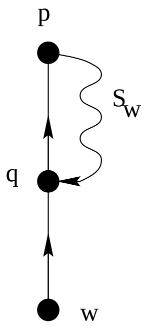


Figure 2.

The model in figure 2. serves as a countermodel for this implication. It is in the characteristic class of P_0 and M_0. One has $w \Vdash p \rhd q$. But clearly $w \nVdash p \rhd q \wedge \Box\neg p$. (Alternatively, the observation that the model is not in the characteristic class of *ILW* would suffice.)

- *ILP₀ W* $\nvdash M_0$.
  This independence result makes essential use of the IL_{set} models. Again a model is provided where at some world P_0 and W hold and M_0 fails. As P_0 semantically implies M_0 using classical frames, we have to move on to IL_{set} models. Figure 3. shows an IL_{set} countermodel for $ILWP_0 \vdash M_0$. In order to show that this is indeed a countermodel, it would be sufficient to note that the frame is in the class of characteristic *ILW* respectively *ILP₀* frames, but not in the class of characteristic *ILM₀* frames. For doing so, one first has to specify the respective classes of characteristic frames.

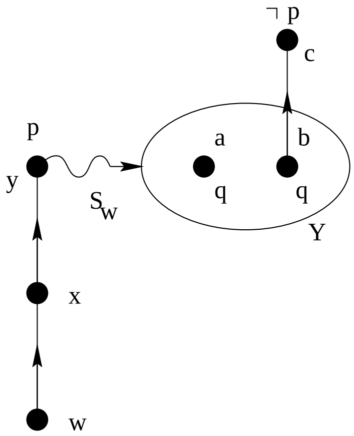


Figure 3.

That however yields rather awkward conditions in this setting. So we will show that in the world w of this specific model, P_0 and W hold as a scheme, whereas M_0 fails. First we will see that P_0 holds as a scheme in w. Suppose therefore $w \Vdash A \rhd \Diamond B$. We must conclude $w \Vdash \Box(A \rhd B)$.
$\Diamond B$ can only hold in x and in b. In the other worlds $\Diamond B$ can not hold because of the absence of R-successors. But both $\{x\}$ and $\{b\}$ can only be accessed with an S_w-transition by x, respectively b itself.

So if $A \rhd \Diamond B$ holds, A can only hold at x or b in which case $\Diamond B$ should hold at the same world where A holds. In particular we see that $\Box\Box\neg A$ holds at w. So definitely $\Box(A \rhd B)$ holds at w. Likewise it is seen that W holds as a scheme at w. We now have to find an instantiation of $\Diamond A \wedge \Box C \rhd B \wedge \Box C$ which does not hold at w. We take $C = A = p$ and $B = q$ to do the job. We have $w \Vdash p \rhd q$ because yS_wY, $y \Vdash p$ and for all elements y' of Y (a and b) one has $y' \Vdash q$. At x, $\Diamond p \wedge \Box p$ holds. But it is impossible to go by an S_w transition to a set of worlds for all of which $q \wedge \Box p$ holds. So indeed M_0 does not hold at w.

QED

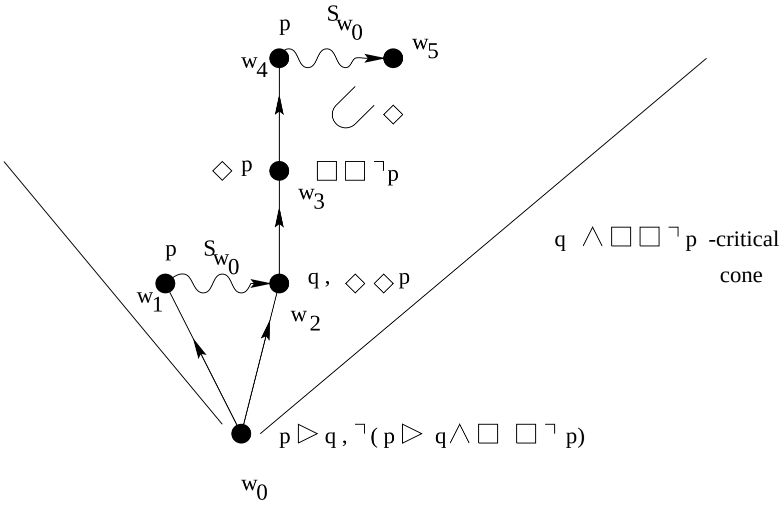
In order to have a complete comparison we note that $ILM \vdash P_0$ and $ILP \vdash P_0$. P_0 holds on every ILM respectively ILP_0 frame. This fact combined with the modal completeness results gives the derivability of P_0 over both ILM and ILP. Of course the odds are for ILW^*P_0 to be an incomplete logic, but we have not been able to prove this due to the lack of a candidate for a valid but underivable principle.

6.4 A relation between M_0 and W

The W-axiom looks quite different from the others. It is easily seen to follow semantically from both M and P. M has been weakened to M_0 and W is no longer derivable over M_0 as was seen in theorem 6.3. M_0 does however semantically imply something similar to W. The frame condition of M_0 excludes the possibility of descending (w.r.t. the R-relation) two worlds with an S_x-relation. This is reflected by a principle W_0, which can be seen as a weakening of the consequent of W.

$$W_0: \quad A \rhd B \to A \rhd B \wedge \Box\Box\neg A.$$

As M_0 is modally complete we must have $ILM_0 \vdash W_0$. Indeed a pure syntactical proof exists. This proof can be extracted from the application of the construction method to this very case. So suppose $ILM_0 \nvdash W_0$, and apply the construction method. We have to construct a model where some instantiation of W_0 does not hold. For that we first consider a maximal ILM_0-consistent set w_0 containing $\neg(p \rhd q \to p \rhd q \wedge \Box\Box\neg p)$. Specifically $p \rhd q \in w_0$ and $\neg(p \rhd q \wedge \Box\Box\neg p) \in w_0$. So in w_0 we have the problem $\neg(p \rhd q \wedge \Box\Box\neg p)$. We solve this problem by introducing a $q \wedge \Box\Box\neg p$ critical successor w_1 of w_0, where p holds. See the figure below. Now we have a deficiency in w_0. So conform the construction method, we add w_2 where q holds.



This w_2 also lies $q \wedge \Box\Box\neg p$-critically above w_0, but as $q \in w_2$ we must have $w_2 \Vdash \Diamond\Diamond p$. So $\Diamond\Diamond p$ is a problem of w_1 (note that $\Diamond\Diamond p \leftrightarrow_{IL} \neg(\Diamond p \rhd \bot)$), and will be eliminated by adding an R-successor w_3 of w_2 such that $w_3 \Vdash \Diamond p$. But also we have $w_3 \Vdash \Box\neg\Diamond p$, i.e. $w_3 \Vdash \Box\Box\neg p$. This was built into lemma 5.3. As $w_3 \Vdash \Diamond p$ we must now introduce the R-successor w_4 of w_3 with $w_4 \Vdash p$. So now we have a deficiency in w_0 w.r.t. w_4. This would be solved by introducing w_4, still in the $q \wedge \Box\Box\neg p$-cone of w_0, such that $w_4 \Vdash q$. We have $w_3 \subset_\Box w_5$, so $\Box\Box\neg p \in w_5$. But as $w_0 \prec_{q \wedge \Box\Box\neg p} w_5$ and $w_5 \Vdash q$, we should have $\neg\Box\Box\neg p \in w_5$. This can not be the case so indeed W_0 must hold. This argument can be captured in a purely syntactical proof.

Suppose $A \rhd B$. ($A \rhd B \in w_2$.) Then one has $A \rhd (B \wedge \Box\Box\neg A) \vee (B \wedge \Diamond\Diamond A)$.(The last disjunct holds at w_2.) But $B \wedge \Diamond\Diamond A \rhd \Diamond A$.(We have $w_3 \Vdash \Diamond A$.), so also $B \wedge \Diamond\Diamond A \rhd \Diamond A \wedge \Box\neg\Diamond A$. (This trick is incorporated in lemma 5.3.) Under the assumption of $A \rhd B$, by M_0 we have $\Diamond A \wedge \Box\Box\neg A \rhd B \wedge \Box\Box\neg A$. (This is the translation of $w_3 \subset_\Box w_5$.) Using transitivity we obtain: $A \rhd B \to B \wedge \Diamond\Diamond A \rhd B \wedge \Box\Box\neg A$. So we can derive $A \rhd B \to A \rhd B \wedge \Box\Box\neg A$ in *ILM₀*.

6.5 The arithmetical validity of P_0

The new principle P_0 is next seen to be arithmetically valid. The argument is due to Albert Visser and can be presented in five basic steps. We will use here standard notation and well-known facts of arithmetization. For a good background one can consult very clear texts like for example [Vis90], [JdJ98]. In the sequel of this paragraph $\Box$ and $\rhd$ will stand again for formalized provability and interpretability respectively. These notions are dependent on the base theory in which they are formalized. In case of possible confusion we will write this base theory by indexing the operator, e.g. $\Box_T$. We have to prove that in any reasonable arithmetical theory $A \rhd \Diamond B \to \Box(A \rhd B)$ is derivable.

For any reasonable arithmetical theory T:

1. Suppose $A \rhd \Diamond B$ and reason within the theory. We thus have $(T + A) \rhd (T + \Diamond B)$.

2. We can now take a finite subtheory T_0 of T which is sufficiently strong. We will not be too precise about this. The actual constraints on T_0 can be distilled from the following argument. It seems natural to demand that T_0 can code the Henkin construction used in the completeness theorem. By this one can obtain a stronger variant of $J5$ as is outlined later on. We have $(T + A) \rhd (T_0 + \Diamond B)$, which turns out to be a Σ_1^0-sentence as T_0 is a finite theory. The Π_2^0-sentence (modulo notational inaccuracies) $\forall y(Ax_{T_0}(y) \to \Box_T(A \to y^J))$ can be replaced by the Σ_1^0-sentence $\Box_T(A \to \tau^J)$, where J is the interpretation used in 1 an τ is the conjunction of all the axioms of T_0.

3. We can thus conclude $\Box_T((T + A) \rhd (T_0 + \Diamond B))$. This follows from 2 and the provable Σ_1^0-completeness of T.

4. The axiom $J5 : \Diamond B \rhd B$ reflects the completeness theorem. As the whole Henkin construction can be coded within T, in T one can prove $(T + Con(T + B)) \rhd (T + B)$. But actually something stronger holds as well. The finite theory T_0 is chosen strong enough to perform the Henkin construction for T. So within T one can prove
$(T_0 + Con(T + B)) \rhd (T + B)$. With necessitation one obtains $\Box_T(T_0 + Con(T + B) \rhd (T + B))$.

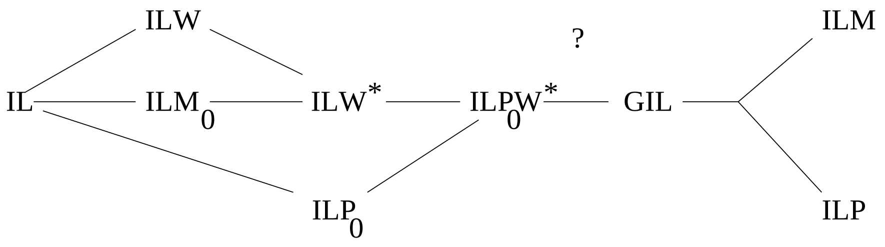
5. The final step consists of combining 3 and 4. The transitivity of $\rhd$ under the $\Box$ yields the desired result, that is $\Box(A \rhd B)$.

7 Concluding

7.1 The new situation

In this paper we have seen the logic *ILM₀* to be modally complete. The decidability is unknown so far. Furthermore a new logic *ILP₀* is introduced which is seen to be modally incomplete. It remains unknown whether ILW^*

is modally complete or not. The author and Dick de Jongh conjecture *ILW** to be modally complete, but the logic ILP_0W^* to be modally incomplete, and to have *GIL* somewhere in between ILP_0W^* and the meet of *ILM* and ILP. *GIL* is assumed to be stronger than ILP_0W^* since this logic is likely to be incomplete and the situation that *GIL* is incomplete is to ghastly to imagine. On the other hand *GIL* cannot be the meet of *ILM* and ILP because this does not have the disjunction property: $\vdash \Box A \vee \Box B \Rightarrow \vdash A$ or $\vdash B$. Theorem 6.3 tells us that the new situation is as depicted below.



7.2 Spin off

- **Modal completeness of *ILW**.**
  The main approach to a modal completeness result is always the same. However for every distinct logic a special ingredient is required to incorporate the frame condition of the added principle into the model construction. One new ingredient used for *ILM$_0$* is given by lemmas 5.5 and 5.6. Another novelty is the local character of the construction method. Instead of defining the model in one blow and defining the S and R relations, we adhere to locally defining all entities to gradually build up the model. Dick de Jongh and Frank Veltman have given a completeness proof of *ILW** in [dJV]. It might be the case that both ingredients of *ILM$_0$* and *ILW* can be combined resulting in a completeness proof of *ILW**. We conjecture *ILW** to be decidable and modally complete. If both completeness proofs can be combined, the decidability will very likely be easier to establish than for *ILM$_0$* by itself.

- **Essentially Σ_1^0-ness.**
  In an article by Dick de Jongh and Duccio Pianigiani a theorem about essentially Σ_1^0-ness is proved. A sentence A in for example the language

of Löbs logic is essentially Σ_1^0 in PA if for any $*$, there is some $\sigma \in \Sigma_1^0$ such that $PA \vdash A^* \leftrightarrow \sigma$.

Theorem 7.1 *(Visser [Vis89], de Jongh [dJD90]) A sentence A in the language of Löbs logic is essentially Σ_1^0 iff A is provably equivalent to some formula of the form $\bigvee_i \Box A_i$.*

Most likely this result can be extended to for example ILM. De Jongh uses the fact that σ is Σ_1^0 in PA iff $PA \vdash \alpha \rhd \beta \to \alpha \wedge \sigma \rhd \beta \wedge \sigma$ for all α, β. Now suppose A is essentially Σ_1^0 but not provably equivalent to some $\bigvee_i \Box A_i$. A model of $p \rhd q$ and $\neg(p \wedge A \rhd q \wedge A)$ is made. It seemed quite difficult to extend this method to ILM. The author and Rosalie Iemhoff have payed some effort to do so. With the construction method presented in this paper it might be possible to reduce the problem to an easier statement.

7.3 Further research

- For the $\Box$-modality we know many readings. We have treated in this paper the interpretation of $\Box$ as the provability predicate. But also other readings are possible. The most prominent (and also the original) reading is that of necessity. Modalities can often have various interpretations. (Think of epistemic logic, temporal logic, etc.) It is always an interesting venture to try to vary the interpretation of a logic in a certain way to then study the purport of the logic under this new interpretation. One obvious variation is to go back to the original inspiration for the semantics of interpretability logics, namely to read the $\rhd$-modality as a conditional in the setting of entailment logics as in Veltman's dissertation. See [Vel85]. As far as we know this has never been done.

- Understanding the relation between Veltman semantics and the arithmetical properties.
  The Veltman semantics have proved to be a very fruitful tool. However, there is no clear understanding about what is the precise relation between Veltman frames and the arithmetics. Neither is there a clear intuitive way of thinking about frame conditions in terms of the intended arithmetic. It might be interesting to investigate if such a clear connection can be described.

- Understanding why certain principles are not complete w.r.t. their corresponding class of characteristic frames.
  When the principle P_0 was found, it was immediately conjectured to be modally incomplete. The grounds for this conjecture were the similarities with an earlier investigated modally incomplete principle $KW4$, and a general intuition. It might be interesting to try to capture this intuition by a general theorem about incompleteness.

- Performing a schematic enclosure of *GIL* as proposed.
  In paragraph 6.2 a general approach is proposed for the enclosure of *GIL*. As far as we know, such a systematic approach has not yet been executed. Within this program various other questions fit in well, like for example can one say something about nice principles in the meet of *ILP* and *ILM*?

8 Acknowledgments

I would like to dedicate this thesis to Hänk Angerman who introduced me to the world of mathematics at the age of twelve. I will not completely blame him for me going into mathematics but he was a great stimulant. Then there is many people how I would like to pay some gratitude. First of all there is Dick de Jongh who was able to support me during all that period when today was mañana but also when mañana all of a sudden was today. I think I have been supervised very well and learned a lot during the production of my thesis. Also Albert Visser has been a great support. The sessions with Frank Veltman, Dick de Jongh and Rosalie Iemhoff proved to be very fruitful. I would like to direct some special word of gratitude to Carlos Areces with whom I shared a room at the department for some time. His many contributions have speeded up enormously the writing of this paper. Also some word of thanks for Johan van Benthem and Anne Kaldewaij with whom I collaborated on giving the first year's course on Logic and Programming. Via this course I got the idea of using invariants while I got payed at the mean time. The ILLC has been very generous to me with all their very useful facilities. (I even got a Christmas box!) Y gracias también a Emma por el apoyo y el cariño durante el desarollo de esta tesis. Then I would like to thank Rosalie and many other people as well. Probably I forget half of the people I should mention. Sorry for that folks. Finally a word of gratitude to my mother not only for her support the last months but also for having introduced me to this wonderful world in which the best of all worlds is at least interpretable.

Contents

References

[AdJH98] C. Areces, D. de Jongh, and E. Hoogland. The interpolation theorem for il and ilp. In *Proceedings of AiML98. Advances in Modal Logic*, Uppsala. Sweden, October 1998. Uppsala University.

[Ben84] J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic. Vol II*, volume 165 of *Synthese Library*, pages 167–248. D. Reidel Publishing Co., Dordrecht, 1984. Extensions of classical logic.

[Ber90] A. Berarducci. The interpretability logic of Peano arithmetic. *Journal of Symbolic Logic*, 55:1059–1089, 1990.

[Boo93] G. Boolos. *The logic of provability.* Cambridge University Press, 1993.

[BV93] A. Berarducci and R. Verbrugge. On the provability logic of bounded arithmetic. *Annals of Pure and Applied Logic*, 61:75–93, 1993.

[Cha91] L. Chagrova. An undecidable problem in correspondence theory. *J. Symbolic Logic*, 56(4):1261–1272, 1991.

[dJD90] D. de Jongh and Pianigiani D. Solution of a problem of david guaspari. Technical report, Universiteit van Amsterdam, 1990.

[dJV] D. de Jongh and F. Veltman. Modal completeness of ILW. Unpublished.

[Göd92] K. Gödel. *On formally undecidable propositions of Principia mathematica and related systems.* Dover Publications Inc., New York, 1992. Translated from the German and with a preface by B. Meltzer, With an introduction by R. B. Braithwaite, Reprint of the 1963 translation.

[Iem98] R. Iemhoff. A modal analysis of some principles of the provability logic of Heyting Arithmetic. In *In proceedings of AiML'98*, Uppsala, 1998.

[JdJ98] G. Japaridze and D. de Jongh. The logic of provability. In S. Buss, editor, *Handbook of proof theory*, pages 475–546. North-Holland Publishing Co., amsterdam edition, 1998.

[JV91] D. de Jongh and A. Visser. Explicit fixed points in interpretability logic. *Studia Logica*, 50:39–50, 1991.

[Löb55] M. H. Löb. Solution of a problem of Leon Henkin. *J. Symb. Logic*, 20:115–118, 1955.

[MVar] M. Marx and Y. Venema. A modal logic of relations. In E. Orlowska, editor, *Memorial Volume for Elena Rasiowa*, Studia Logica Library. Kluwer Academic Publishers, Dordrecht, to appear.

[Pet90] P. Petkov, editor. *Mathematical logic, Proceedings of the Heyting 1988 summer school in Varna, Bulgaria.* Plenum Press, Boston, 1990.

[Sha88] V. Shavrukov. The logic of relative interpretability over Peano arithmetic (in Russian). Technical Report Report No.5, Stekhlov Mathematical Institute, Moscow, 1988.

[Sha94] V. Shavrukov. A smart child of Peano's. *The Notre Dame Journal of Formal Logic*, 35:161–185, 1994.

[Šve91] V. Švejdar. Some independence results in interpretability logic. *Studia Logica*, 50(1):29–38, 1991.

[TMR53] A. Tarski, A. Mostowski, and R. Robinson. *Undecidable theories.* North–Holland, Amsterdam, 1953.

[Vel85] F. Veltman. *Logic for conditionals.* PhD thesis, Department of Philosophy, University of Amsterdam, 1985.

[Ver] R. Verbrugge. Verzamelingen-Veltman frames en modellen. Unpublished.

[Vis84] A. Visser. The provability logics of recursively enumerable theories extending Peano arithmetic at arbitrary theories extending Peano arithmetic. *Journal of Philosophical Logic*, 13:97–113, 1984.

[Vis89] A. Visser. Peano's smart children: A provability logical study of systems with built-in consistency. *Notre Dame Journal of Formal Logic*, 30:161–196, 1989.

[Vis90] A. Visser. Interpretability logic. In *[Pet90]*, pages 175–209, 1990.

[Vis91] A. Visser. The formalization of interpretability. *Studia Logica*, 51:81–105, 1991.

[Vis94] A. Visser. *Propositional combinations of* Σ*–sentences in Heyting's Arithmetic.* Logic Group Preprint Series 117, 1994.

[Vis97] A. Visser. An overview of interpretability logic. In M. Kracht, M. de Rijke, and H. Wansing, editors, *Advances in modal logic '96*, pages 307–359. CSLI Publications, Stanford, CA, 1997.